



Vaasan yliopisto  
UNIVERSITY OF VAASA

Tuukka Hyytiäinen

## **Kustannusten ennustaminen TimeGPT-mallilla**

Tekniikan ja innovaatiojohtamisen yksikkö  
Kauppatieteiden pro gradu- tutkielma  
Tietojärjestelmätiede

Vaasa 2025

---

**VAASAN YLIOPISTO****Tekniikan ja innovaatiojohtamisen yksikkö**

<b>Tekijä:</b>	Tuukka Hyytiäinen		
<b>Tutkielman nimi:</b>	Kustannusten ennustaminen TimeGPT-mallilla		
<b>Tutkinto:</b>	Kauppatieteiden maisteri		
<b>Oppiaine:</b>	Tietojärjestelmätiede		
<b>Työn ohjaaja:</b>	Tomi Pasanen, Mikko Ranta		
<b>Valmistumisvuosi:</b>	2025	<b>Sivumäärä:</b>	107

---

**TIIVISTELMÄ:**

Tässä pro gradu -tutkielmassa syvennytään TimeGPT-tekoälymallin soveltuvuuteen kohdeorganisaation kustannusten ennustamisessa. Tarkka ennustaminen on kohdeorganisaatiolle tärkeää resurssien kohdentamisen ja strategisen päätöksenteon tueksi, ja perinteiset menetelmät ovat osoittautuneet osin riittämättömiksi monimutkaisen datan käsittelyssä. Tutkimuksen tavoitteena oli arvioida TimeGPT:n suorituskykyä ja tarkkuutta verrattuna vakiintuneisiin aikasarjamenetelmiin (ETS, SARIMA, Prophet) sekä organisaation aiempiin ennustekäytäntöihin. Lisäksi selvitettiin ennustajien käyttäjäkokemuksia ja teknologian hyväksyntään vaikuttavia tekijöitä uuden tekoälyavusteisen työkalun käyttöönotossa.

Tutkimus toteutettiin mixed methods -lähestymistavalla, hyödyntäen kohdeorganisaation kuukausittaista kustannusdataa (1/2020–10/2024) eri hierarkia- ja kompleksisuustasoilla. Kvantitatiivisessa analyysissä mallien ennustetarkkuutta (mm. MAE, MAPE) ja laskentanopeutta verrattiin, ja erojen tilastollista merkitsevyyttä arvioitiin Diebold-Mariano-testillä. Kvalitatiivinen aineisto kerättiin ennustajille (N≈118) suunnatulla kyselyllä, keskittyen Power BI -raporttiin integroidun TimeGPT-ennustetyökalun koettuun helppokäyttöisyyteen, hyödyllisyyteen ja luotettavuuteen (TAM/UTAUT-viitekehysten kautta).

Kvantitatiiviset tulokset osoittivat TimeGPT:n olevan laskennallisesti ylivoimaisen nopea. Se tuotti myös tarkempia ennusteita kuin vertailumallit erityisesti käsiteltäessä laajoja, monimutkaisia ja osin puutteellisia data-aineistoja. Mallin hienosäätö organisaation omalla datalla paransi tarkkuutta verrattuna zero-shot-lähtötasoon, joskin lisähyöty väheni noin 30–60 hienosäätöaskeleen jälkeen. TimeGPT ylitti tarkkuudessa myös aiemmat manuaaliset ennusteet. Kvalitatiiviset havainnot toivat esiin merkittävän kontrastin: Power BI -raportin visualisointi- ja porautumisominaisuudet koettiin erittäin hyödyllisiksi ja helppokäyttöisiksi, mutta itse TimeGPT-mallin "musta laatikko" -luonne ja toimintalogiikan epäselvyys herättivät laajaa epäluottamusta ja jarruttivat sen aktiivista hyödyntämistä (vain 50% kyselyn vastaajista käytti sitä).

Johtopäätöksenä TimeGPT tarjoaa kohdeorganisaatiolle huomattavaa teknistä potentiaalia ennusteprosessien nopeuttamiseen ja tarkentamiseen, erityisesti haastavissa dataolosuhteissa. Sen täysimääräinen ja onnistunut käyttöönotto vaatii kuitenkin kriittisesti panostuksia mallin toiminnan selittämiseen, läpinäkyvyyden lisäämiseen ja käyttäjien koulutukseen luottamuksen ja hyväksynnän saavuttamiseksi. Tämä tutkimus tarjoaa kohdeorganisaatiolle käytännön suosituksia ja edistää ymmärrystä tekoälyn perustamallien integroinnista organisaatioiden talousohjaukseen.

---

**AVAINSANAT:** Aikasarjaennustaminen, Tekoälymallit, TimeGPT, Kustannusten ennustaminen, Transformer-arkkitehtuuri, Ennustemallien vertailu

---

**UNIVERSITY OF VAASA****School of Technology and Innovations**

**Author:** Tuukka Hyytiäinen  
**Title of the thesis:** Cost forecasting using the TimeGPT model  
**Degree:** Master of Science in Economics and Business Administration  
**Programme:** Master's Degree Programme in Information Systems  
**Supervisor:** Tomi Pasanen, Mikko Ranta  
**Year:** 2025                      **Pages:** 107

---

**ABSTRACT:**

This Master's thesis delves into the applicability of the TimeGPT AI model for cost forecasting within the target organization. Accurate forecasting is vital for the target organization to support resource allocation and strategic decision-making, and traditional methods have proven partially insufficient in handling complex data. The objective of the study was to evaluate the performance and accuracy of TimeGPT compared to established time series methods (ETS, SARIMA, Prophet) and the organization's previous forecasting practices. Additionally, the study investigated forecasters' user experiences and factors influencing technology acceptance during the adoption of the new AI-assisted tool.

The research was conducted using a mixed-methods approach, utilizing the target organization's monthly cost data (1/2020–10/2024) at various hierarchy and complexity levels. In the quantitative analysis, the models' forecasting accuracy (incl. MAE, MAPE) and computational speed were compared, and the statistical significance of the differences was assessed using the Diebold-Mariano test. Qualitative data was collected through a survey directed at forecasters (N≈118), focusing on the perceived ease of use, usefulness, and reliability of the TimeGPT forecasting tool integrated into a Power BI report (through TAM/UTAUT frameworks).

The quantitative results showed TimeGPT to be computationally superior in speed. It also produced more accurate forecasts than the comparison models, especially when handling large, complex, and partially incomplete datasets. Fine-tuning the model with the organization's own data improved accuracy compared to the zero-shot baseline, although the additional benefit diminished after approximately 30–60 fine-tuning steps. TimeGPT also surpassed previous manual forecasts in accuracy. The qualitative findings revealed a significant contrast: the visualization and drill-down features of the Power BI report were perceived as highly useful and easy to use, but the "black box" nature of the TimeGPT model itself and the lack of clarity in its operating logic generated widespread distrust and hindered its active utilization (only 50% of survey respondents used it).

In conclusion, TimeGPT offers the target organization considerable technical potential for accelerating and refining forecasting processes, especially under challenging data conditions. However, its full and successful adoption critically requires, in addition to technical performance, investments in explaining the model's operation, increasing transparency, and user training to achieve trust and acceptance. This study provides practical recommendations for the target organization and advances the understanding of integrating AI foundation models into organizational financial management.

---

**KEYWORDS:** Time series forecasting, AI models, TimeGPT, Cost forecasting, Transformer architecture, Comparison of forecasting models

## Sisällys

1	Johdanto	9
1.1	Tausta ja motivaatio	9
1.2	Tutkimusongelma	10
1.3	Tutkimuksen tavoitteet	11
1.4	Tutkimuskysymykset	11
1.5	Tutkimuksen merkitys	12
2	Aikasarjaennustamisen menetelmät ja arviointi	14
2.1	Aikasarjaennustamisen perusteet	14
2.2	Perinteiset ennustemallit	15
2.2.1	Ekspontiaalinen tasoitus (ETS)	15
2.2.2	SARIMA-malli	16
2.2.3	Prophet	18
2.3	Tekoäly ja koneoppiminen aikasarjaennustamisessa	19
2.3.1	Transformer-arkkitehtuuri	19
2.3.2	TimeGPT-malli	20
2.3.3	Koneoppimisen sovellukset aikasarjaennustamisessa	22
2.4	Ennustemallien vertailu ja arviointi	24
2.4.1	Ennustemallien suorituskyvyn mittarit	26
2.4.2	Tilastolliset testit ennustemallien vertailussa	28
2.5	Ennustajien kokemukset ja teknologian omaksuminen	29
2.5.1	Kvalitatiivinen tutkimus teknologian hyväksymisestä	29
2.5.2	Ennustajien suhtautuminen tekoälypohjaisiin ratkaisuihin	30
2.6	Aikaisemmat tutkimukset	32
2.6.1	TimeGPT:n soveltaminen eri aloilla	32
2.6.2	Suuret kielimallit aikasarjaennustamisessa	33
2.6.3	Hybridimallit	34
2.6.4	Teknologian omaksuminen ja käyttäjäkokemukset	34
2.6.5	Yhteenvedo aikaisemmista tutkimuksista	35
2.7	Tutkimusaukkojen tunnistaminen	36

3	Tutkimusmenetelmät ja empiirinen osuus	38
3.1	Aineiston keruu	38
3.1.1	Kyselylomakkeen toteutus ja sisältö	40
3.2	Datan esikäsittely	41
3.3	Ennustemallit ja niiden toteutus	43
3.4	Data-analyysi	46
3.5	Tutkimusmenetelmiin ja dataan liittyvät eettiset näkökulmat	50
4	Tulokset	52
4.1	Ennustemallien suorituskyky	52
4.1.1	Ennustemallien väliset tilastolliset erot	52
4.1.2	Ennusteet eri tasoilla ja ennustemalleilla	53
4.1.3	TimeGPT:n hienosäädön askelmäärän vaikutus	57
4.1.4	Ennustemallien ajoaikojen vertailu	60
4.1.5	Vertailu ennustajien aiempaan ennusteeseen	61
4.2	Kvalitatiiviset havainnot	63
4.2.1	Kokemusten analyysi	63
4.2.2	Ennustajien kokemukset	66
5	Johtopäätökset ja pohdinta	69
5.1	Analyysi	69
5.1.1	Suorituskyky suhteessa datan monimutkaisuuteen	69
5.1.2	Hienosäädön vaikutus ja optimaalinen taso	70
5.1.3	Laskennallinen tehokkuus ja käytännön näkökohdat	70
5.1.4	Vertailu aiempiin ennustekäytäntöihin	70
5.1.5	Käyttäjäkokemukset ja teknologian hyväksyntä	71
5.1.6	Yhteenvedo analyysistä	71
5.2	Tutkimuksen merkitys	72
5.2.1	Käytännön merkitys kohdeorganisaatiolle	72
5.2.2	Tieteellinen merkitys	74
5.3	Tutkimuksen rajoitukset	75
5.4	Jatkotutkimusaiheet	76

Lähteet	78
Liitteet	84
Liite 1. DM-HLN testi	84
Liite 2. Ajallisen cross-validationin toteutus MAE-mittarilla	86
Liite 3. SARIMA ennuste	89
Liite 4. Prophet ennuste	93
Liite 5. ETS ennuste	95
Liite 6. TimeGPT zero-shot ennuste	97
Liite 7. TimeGPT hienosäädetty ennuste	99
Liite 8. Kysely ennustajille	101
Liite 9. Kyselyn vastaukset	104

## Kuvat

**Kuva 1.** Power BI ennustetyökalu

**Kuva 2.** Kaikki yhtiöt ja alin taso

**Kuva 3.** Yksi yhtiö ja kaikki tasot

**Kuva 4.** Kaikki yhtiöt ja kaikki tasot

**Kuva 5.** TimeGPT ennusteiden kk-MAE ja kokonais-MAE eri hienosäädön askelmäärillä

**Kuva 6.** TimeGPT ennusteet eri hienosäädön askelmäärillä

**Kuva 7.** TimeGPT-ennusteet ja ennustajien tekemä ennuste

## Taulukot

**Taulukko 1.** SARIMA-mallin komponentit.

**Taulukko 2.** Yhteenveto neljästä tutkitusta ennustemallista.

**Taulukko 3.** Ennustemallien mittarit.

**Taulukko 4.** Diebold-Mariano-testin korjatut p-arvot.

**Taulukko 5.** Kaikki yhtiöt ja alin taso

**Taulukko 6.** Yksi yhtiö ja kaikki tasot

**Taulukko 7.** Kaikki yhtiöt ja kaikki tasot

**Taulukko 8.** TimeGPT ennusteet eri hienosäädön askelmäärillä

**Taulukko 9.** Ennustemallien ajojen kesto (sekuntia)

**Taulukko 10.** TimeGPT-ennusteet ja ennustajien tekemä ennuste

## Lyhenteet

AIC	Akaike Information Criterion
API	Application Programming Interface
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
BI	Business Intelligence
BIC	Bayesian Information Criterion
CNN	Convolutional Neural Network
CSV	Comma-Separated Values

DM	Diebold-Mariano (testi)
ESP	Electrical Submersible Pump
ETS	Exponential Smoothing
GRU	Gated Recurrent Unit
HLN	Harvey-Leybourne-Newbold (korjaus)
K1	Konsernin ensimmäinen ennustekierros
LLM	Large Language Model
LLMTIME	Large Language Model for Time Series
LSTM	Long Short-Term Memory
LSTPrompt	Long Short-Term Prompting
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC-EWS	Mayo Clinic – Early Warning Score
MSE	Mean Squared Error
N-BEATS	Neural Basis Expansion for Time-Series
PEOU	Perceived Ease of Use
Prophet	(Meta/Facebook-mallin nimi)
PU	Perceived Usefulness
RMSE	Root Mean Squared Error
SARIMA	Seasonal ARIMA
SCADA	Supervisory Control And Data Acquisition
SDG	Sustainable Development Goals
TAM	Technology Acceptance Model
TimeGPT	Time-series Generative Pre-trained Transformer (mallin nimi)
Time-LLM	Time-series-adapted LLM
TimelyGPT	(mallin nimi)
TSLANet	Time-Series Learning Attention Network
UTAUT	Unified Theory of Acceptance and Use of Technology
xPos	Extrapolatable Positional Embedding

# 1 Johdanto

## 1.1 Tausta ja motivaatio

Liikenteenohjauksen ja -hallinnan alalla tarkkojen kustannusennusteiden laatiminen on saanut keskeisen aseman organisaatioiden pyrkiessä optimoimaan resurssien käyttöä ja tukemaan päätöksentekoa. Erityisesti suurissa konserneissa, kuten kohdeorganisaatiossa, budjetoinnin onnistuminen niin tytäryhtiöiden kuin emoyhtiön tasolla edellyttää luotettavaa tietoa tulevista kustannuksista. Perinteiset tilastolliset ennustemallit (kuten SARIMA ja ETS) sekä uudemmat ratkaisut, kuten Prophet, voivat kuitenkin osoittautua tehottomiksi tai hankaliksi ylläpitää, erityisesti silloin kun käsiteltävä data on kooltaan suurta, monimutkaista, sisältää puutteita tai merkittäviä ajallisia riippuvuuksia.

Viime vuosina tekoälypohjaiset ennustemallit, erityisesti generatiivisiin transformer-arkkitehtuureihin perustuvat ratkaisut, ovat osoittaneet lupaavia tuloksia monilla toimialoilla. Esimerkkinä tästä on Nixtlan kehittämä TimeGPT, joka edustaa ensimmäisiä yleiskäyttöisiä aikasarjaennustamisen malleja. Mallin kyky tuottaa tarkkoja ennusteita ilman laajaa uudelleen koulutusta (zero-shot-ominaisuus) sekä mahdollisuus hienosäätää (fine-tuning) mallia organisaatiokohtaisella datalla luovat merkittäviä mahdollisuuksia tarkempaan kustannusennustamiseen.

Tässä tutkimuksessa keskitytään vain kohdeorganisaation kustannusdatan ennustamiseen. Tutkimus ei sisällä tulo- tai kassavirtatarkastelua, eikä se kata muiden liikenteenohjausalalla toimivien yritysten vastaavia aineistoja. Tutkimuksen aikajänne (1/2020–10/2024) on valittu rajoitetusti, sillä se kattaa sekä budjetoinnin että resurssien käytön kannalta olennaiset sekä eheät historia- ja ennustejaksot. Tutkimuksen tarkkuutta lisätäkseen ulkopuoliset ja makrotason muuttujat, kuten yleinen taloustilanne tai liikenedatan määrät, on rajattu mallista pois. Tämän avulla mahdollistetaan TimeGPT-mallin ja perinteisten mallien suoraviivainen vertailu organisaation sisäisen datan pohjalta. Rajaaminen on käytännöllinen valinta, mutta samalla se tarkoittaa, että ulkoisten tekijöiden

vaikutuksia kustannuksiin ei analysoida tässä työssä. Tutkimuksen tulosten sovelletta-  
vuus muiden yritysten tai toimialojen konteksteihin on rajallista ilman vastaavaa data-  
aineistoa ja organisaatiokohtaista ymmärrystä. Jotta tarkempia johtopäätöksiä voitaisiin  
tehdä, vaaditaan vähintään yhtä laajaa ja yksityiskohtaista aineistoa myös muista läh-  
teistä.

Tässä tutkimuksessa tarkastellaan TimeGPT-mallin soveltuvuutta kustannusten ennusta-  
miseen kohdeorganisaatiossa. Samalla analysoidaan ennustajien asenteita tekoälyavus-  
teista ennustamisprosessia kohtaan sekä arvioidaan, miten tällainen ratkaisu voisi vai-  
kuttaa organisaation toimintaan. Tutkimuksessa yhdistyy kvantitatiivinen mallivertailu ja  
kvalitatiivinen kyselyaineisto, mikä mahdollistaa sekä mallin teknisen suorituskyvyn että  
sen käytettävyyden ja hyödyllisyyden kattavan tarkastelun loppukäyttäjien näkökulmasta.

## 1.2 Tutkimusongelma

Kohdeorganisaation kustannusrakenteen ennustaminen on perinteisesti pohjautunut  
ennustajien tekemään historiatiedon analysointiin sekä ennustajien omiin arvioihin.  
Konsernin monimutkainen kulurakenne ja datan mahdolliset puutteet voivat kuitenkin  
heikentää näiden menetelmien tuottamien ennusteiden luotettavuutta. Tutkimuson-  
gelma tiivistyy seuraavasti:

*Miten TimeGPT-ennustemallia voidaan hyödyntää hyödyntää kohdeorganisaation  
kustannusennustamisessa, ja kuinka tehokas se on verrattuna perinteisiin mallei-  
hin sekä konsernin aiempiin ennustekäytäntöihin?*

Tutkimusongelmaan liittyy myös jatkokysymyksiä, jotka käsittelevät uuden tekoälypoh-  
jaisen mallin vaikutusta ennustajien työhön. Näihin lukeutuvat erityisesti mallin koettu  
helppokäyttöisyys ja sen luotettavuus osana organisaation päivittäisiä prosesseja. Näin  
voidaan arvioida paitsi mallin teknistä suorituskykyä, myös sen käytettävyyttä ja arvoa  
organisaation näkökulmasta.

### 1.3 Tutkimuksen tavoitteet

Tutkimuksen päätavoitteena on arvioida TimeGPT-mallin suorituskykyä ja käytettävyyttä kustannusten ennustamisessa kohdeorganisaatiossa. Tämä tavoite jakautuu kolmeen keskeiseen osa-alueeseen:

1. **Ennustetarkkuuden vertailu:** TimeGPT:n ennustetarkkuutta verrataan perinteisiin ennustemalleihin (kuten ETS, SARIMA ja Prophet) sekä kohdeorganisaation aiempiin omiin ennusteisiin. Vertailu kohdistuu eri hierarkiatasoihin: tilitasolle, vastuualuekohtaiselle tasolle sekä koko konsernin tasolle. Lisäksi analysoidaan mallin suorituskykyä erilaisilla data-aineiston laajuuksilla.
2. **Käyttäjäkokemus ja teknologian hyväksyntä:** Tutkimuksessa selvitetään, miten ennustajat kokevat TimeGPT-mallin käytön, millaisia hyötyjä ja haasteita he tunnistavat ja miten malli vaikuttaa heidän työprosesseihinsa. Tähän sisältyy tarkastelu teknologian hyväksymisen näkökulmasta hyödyntäen malleja TAM (Technology Acceptance Model) ja UTAUT (Unified Theory of Acceptance and Use of Technology).
3. **Suosituksien ja jatkokehityksen:** Lopuksi tutkimus antaa suosituksia TimeGPT-mallin hyödyntämisestä kustannusennustamisessa. Lisäksi tunnistetaan jatkokehityskohteita, joiden avulla mallia voidaan edelleen räätälöidä tukemaan paremmin liikenteenohjauksen ja -hallinnan tarpeita.

### 1.4 Tutkimuskysymykset

Tutkimus pyrkii vastaamaan seuraaviin kysymyksiin:

1. **Kuinka tarkasti TimeGPT-malli ennustaa kohdeorganisaation kustannukset verrattuna perinteisiin ennustemalleihin ja konsernin omiin ennusteisiin?**

Tarkastellaan mallin suorituskykyä eri hierarkiatasoilla ja data-aineistojen laajuuksilla sekä analysoidaan, kuinka hyvin se vastaa konsernin ennustetarpeisiin.

**2. Mitkä ovat TimeGPT-mallin keskeiset vahvuudet ja heikkoudet kustannusten ennustamisessa?**

Arvioidaan mallin tarkkuuden lisäksi sen soveltuvuutta monimutkaisten ja puutteellisten tietoaaineistojen käsittelyyn.

**3. Miten ennustajat kokevat tekoälypohjaisen TimeGPT-mallin käytön, ja millaisia parannusehdotuksia heillä on ennustetyökalun suhteen?**

Selvitellään ennustajien kokemuksia, teknologian hyväksymistä sekä työnkulkuun liittyviä muutoksia.

**4. Miten TimeGPT-mallia voidaan kehittää edelleen vastaamaan paremmin kohdeorganisaation ennustetarpeita ja käytäntöjä?**

Tunnistetaan kehityskohteita, joiden avulla mallia voidaan mukauttaa entistä paremmin tukemaan kohdeorganisaation operatiivisia ja strategisia tavoitteita.

## **1.5 Tutkimuksen merkitys**

Tarkemmat kustannusennusteet ovat kohdeorganisaatiolle merkittävä väline toiminnan tehokkuuden ja resurssienhallinnan parantamiseen. Virheellinen tai puutteellinen budjetointi voi johtaa yllättäviin lisäkustannuksiin, resurssipulaan tai vaihtoehtoisesti käyttämättömiin resursseihin. Luotettavampien ennusteiden avulla organisaatio pystyy suunnittelemaan ja toteuttamaan hankintoja sekä investointeja oikea-aikaisesti, välttämään päällekkäisiä kustannuksia ja kohdentamaan henkilöstöresursseja entistä tehokkaammin. Käytännössä tämä näkyy esimerkiksi paremmin ajoitetuissa huolto- ja kunnossapitotoimissa sekä resurssien ohjaamisessa tarkoituksenmukaisiin prosesseihin.

Tarkemmat ennusteet tukevat myös kohdeorganisaation tuottamien palveluiden laadun ja turvallisuuden kehittämistä. Oikea-aikaisesti ennakoitujen kustannukset auttavat välttämään säästötoimenpiteitä, jotka voisivat vaarantaa palvelutasoa. Tämä mahdollistaa liikenteen sujuvuuden, asiakastyytyväisyyden ja konsernin strategisen päätöksenteon tu-

kemisen entistä paremmin. Tässä yhteydessä TimeGPT:n kaltaiset tekoälyratkaisut tarjoavat arvokasta tukea kohdeorganisaation talous- ja resurssisuunnittelun kehittämisessä.

Tämä tutkimus tuottaa uutta tietoa tekoälypohjaisten ennustemallien soveltamisesta kustannusennusteisiin liikenteenohjauksen kontekstissa. Käytännön tasolla tutkimus tarjoaa kohdeorganisaatiolle konkreettisia ohjeita ja suosituksia TimeGPT-mallin käyttöön, mikä voi tehostaa ennusteprosesseja ja parantaa ennustetarkkuutta konsernin eri hierarkiatasoilla.

Samalla tutkimus edistää akateemista keskustelua tekoälymallien hyödyntämisestä taloudellisissa ennusteissa, erityisesti monimutkaisten datarakenteiden ja organisaation sisäisten prosessien yhteydessä. Kvalitatiiviset havainnot tarjoavat syvällisemmän käsityksen siitä, miten ennustajat kokevat tekoälyteknologian käytön. Tutkimus voi toimia lähtökohtana myös muille projekteille, joissa pureudutaan tekoälypohjaisten mallien käytettävyyden ja skaalautuvuuskykyyn eri yritys ympäristöissä ja liiketoimintaprosesseissa.

## 2 Aikasarjaennustamisen menetelmät ja arviointi

### 2.1 Aikasarjaennustamisen perusteet

Aikasarjaennustaminen on keskeinen osa monia sovellusalueita, kuten taloutta, energiaa ja terveydenhuoltoa, joissa ennusteet ohjaavat päätöksentekoa ja resurssien kohdentamista. Aikasarja-analyysin ytimessä on ajallisesti järjestetyn datan mallintaminen, jonka tavoitteena on ymmärtää menneisyyden trendejä ja vaihteluita sekä ennustaa tulevia arvoja. Tämä osio esittelee aikasarjaennustamisen keskeisiä menetelmiä ja lähestymistapoja perustuen alan keskeisiin julkaisuihin.

Klassiset tilastolliset menetelmät, kuten Autoregressive Integrated Moving Average (ARIMA), ovat perinteisesti hallinneet aikasarja-analyysiä. Näiden mallien lähtökohtana on datan stationaarisuus, ja ne toimivat erityisesti tilanteissa, joissa sarjojen trendit ja kausivaihtelut voidaan selkeästi tunnistaa ja mallintaa (Box, Jenkins, Reinsel ja Ljung, 2015). ARIMA-mallien etuna on niiden kyky tuottaa yksinkertaisia ja selkeitä ennusteita yhdistämällä autoregressiiviset ja liukuvan keskiarvon komponentit. Mallit ovat tehokkaita lyhyen aikavälin ennusteissa, mutta niiden sovellettavuus heikkenee datan monimutkaisuuden ja pitkän aikavälin riippuvuuksien kasvaessa.

Eksponentiaalisen tasoituksen mallit, kuten Holt-Winters-menetelmä, tarjoavat tehokkaan ratkaisun kausiluonteisten aikasarjojen analyysiin. Näissä menetelmissä korostuvat uusimpien havaintojen painoarvo ennusteiden tuottamisessa, mikä tekee niistä erityisen sopivia muuttuvan datan käsittelyyn. Hyndman ja Athanasopoulos (2018) korostavat eksponentiaalisen tasoituksen roolia erityisesti liiketoiminnallisessa suunnittelussa lyhyen ja keskipitkän aikavälin ennusteissa.

Monimutkaisten ja pitkän aikavälin riippuvuuksien mallintaminen on aikasarjaennustamisen keskeinen haaste. Perinteiset menetelmät, kuten Autoregressiiviset mallit (AR) ja ARIMA, soveltuvat hyvin lineaarisiin ja lyhyen aikavälin riippuvuuksiin, mutta niiden suo-

rituskyky heikkenee monimutkaisten, ei-lineaaristen ja pitkien aikavälien kausaalisuhteiden mallintamisessa. Tällaisissa tilanteissa tarvitaan menetelmiä, jotka kykenevät erottamaan signaalin ja melun sekä havaitsemaan piileviä syy-seuraussuhteita.

Vaikka klassiset tilastolliset mallit ovat tehokkaita tietyissä sovelluksissa, niiden suorituskyky voi jäädä rajalliseksi, kun data on ei-stationaarista tai sisältää epälineaarisia riippuvuuksia. Uudemmat lähestymistavat, kuten syväoppimismenetelmien integrointi, tarjoavat parannuksia tällaisissa tilanteissa. Chatfield (2003) painottaa koneoppimisen ja perinteisten tilastollisten menetelmien yhdistämisen etuja erityisesti silloin, kun data sisältää sekä lineaarisia että ei-lineaarisia komponentteja.

Aikasarjaennustamisen sovellusalueet ulottuvat laajalle. Esimerkiksi taloudessa aikasarjamalleja käytetään ennustamaan osakemarkkinoiden liikkeitä, kun taas energianhallinnassa ne optimoivat resurssien käyttöä. Chatfield (2003) korostaa, että aikasarjojen analyysin ja ennustamisen peruseräotteiden ymmärtäminen on keskeistä päätöksenteossa. Tulevaisuudessa tutkimus keskittyyneen menetelmiin, jotka yhdistävät perinteisten mallien selkeyden ja syväoppimisen joustavuuden.

## **2.2 Perinteiset ennustemallit**

### **2.2.1 Eksponentiaalinen tasoitus (ETS)**

Eksponentiaalinen tasoitus (ETS) on keskeinen menetelmä aikasarjaennustamisessa, erityisesti tilanteissa, joissa data sisältää trendin ja kausivaihtelua. ETS-mallit painottavat tuoreimpia havaintoja voimakkaammin kuin vanhempia, mikä tekee niistä herkkiä uusille muutoksille ja erityisen sopivia lyhyen aikavälin ennusteisiin.

Hyndman ja Athanasopoulos (2018) esittelevät ETS-mallien kolme pääkomponenttia: tason (level), trendin (trend) ja kausivaihtelun (seasonality). Näiden komponenttien yhdistelmät muodostavat erilaisia ETS-malleja, kuten additiivisia ja multiplikatiivisia malleja,

jotka valitaan datan ominaispiirteiden perusteella. Esimerkiksi additiivinen malli soveltuu tilanteisiin, joissa kausivaihtelun suuruus pysyy vakiona ajan myötä, kun taas multiplikatiivinen malli on tarkoituksenmukainen, jos kausivaihtelun voimakkuus vaihtelee suhteessa trendiin.

ETS-mallien etuna on niiden kyky sopeutua nopeasti muuttuviin olosuhteisiin, mikä tekee niistä erityisen hyödyllisiä liiketoiminnan suunnittelussa ja varastonhallinnassa. Hyndman ja Athanasopoulos (2018) korostavat, että ETS-mallit ovat tehokkaita erityisesti silloin, kun data ei noudata selkeää kaavaa tai sisältää satunnaista vaihtelua.

Box, Jenkins, Reinsel ja Ljung (2015) käsittelevät eksponentiaalisen tasoituksen menetelmiä osana laajempaa aikasarja-analyysin kokonaisuutta. Heidän mukaansa ETS-mallit tarjoavat yksinkertaisen mutta tehokkaan tavan ennustaa tulevia arvoja erityisesti silloin, kun datassa esiintyy selkeä trendi tai kausivaihtelua. Lisäksi Chatfield (2003) huomauttaa, että eksponentiaalinen tasoitus on erityisen hyödyllinen tilanteissa, joissa data sisältää satunnaista vaihtelua. Menetelmä tasoittaa lyhyen aikavälin vaihtelut ja tuo esiin pitkän aikavälin trendit. Tämä tekee ETS-malleista soveltuvia monenlaisiin sovelluksiin, kuten taloudellisten indikaattorien ennustamiseen ja tuotannon suunnitteluun.

### 2.2.2 SARIMA-malli

Seasonal Autoregressive Integrated Moving Average (SARIMA) -malli on laajennus perinteisestä ARIMA-mallista, joka huomioi aikasarjojen kausiluonteiset komponentit. SARIMA-malli yhdistää autoregressiiviset (AR), integroivat (I) ja liukuvan keskiarvon (MA) osat sekä lisää kausiluonteiset komponentit. Tämä tekee mallista erityisen sopivan aikasarjoille, joissa esiintyy säännöllistä kausivaihtelua, kuten kuukausittaisessa myynnissä, energiankulutuksessa tai säätilastojen analyysissä. Taulukossa 1 SARIMA-mallin komponenttien rakenne:  $ARIMA(p,d,q)(P,D,Q)[m]$ .

**Taulukko 1.** SARIMA-mallin komponentit.

<b>Komponentti</b>	<b>Selitys</b>
<b>p</b>	Autoregressiivisen osan järjestys, joka määrittää edellisten havaintojen vaikutuksen nykyiseen arvoon.
<b>d</b>	Differoinnin aste, joka poistaa trendin datasta.
<b>q</b>	Liukuvan keskiarvon osan järjestys, joka tasoittaa melua ja muita satunnaisia vaihteluita.
<b>P, D, Q</b>	Kausiluonteiset ARIMA-komponentit, jotka käsittelevät datan kausivaihtelua.
<b>m</b>	Kausivaihtelun jakson pituus, esimerkiksi kuukausittaisessa aikasarjassa m on 12.

Box, Jenkins, Reinsel ja Ljung (2015) kuvaavat SARIMA-mallin keskeisenä etuna sen kykyä yhdistää trendin ja kausivaihtelun mallintaminen yhtenäiseen kehykseen. Tämä tekee mallista erityisen hyödyllisen datalle, jossa esiintyy monimutkaisia riippuvuuksia. Esimerkiksi talouden tai liiketoiminnan analyysissä SARIMA voi auttaa ennustamaan myyntitrendejä, joissa yhdistyvät vuosittainen kausivaihtelu ja pitkän aikavälin kasvutrendi.

Hyndman ja Athanasopoulos (2018) korostavat SARIMA-mallien laajaa sovellettavuutta, mainiten erityisesti taloudellisten indikaattorien, sääennusteiden ja logistiikan kysynnänhallinnan ennusteet. Heidän mukaansa mallin suurin vahvuus on sen joustavuus: kun kausivaihtelun jakso on tiedossa, SARIMA-mallit pystyvät sopeutumaan sekä lyhyen että pitkän aikavälin muutoksiin.

Shumway ja Stoffer (2017) nostavat esiin SARIMA-mallin kyvyn hallita sekä stationaarisia että ei-stationaarisia komponentteja. Tämä on erityisen tärkeää sovelluksissa, joissa sekä trendit että kausivaihtelu voivat muuttua ajan myötä. He mainitsevat myös SARIMA-mallin soveltuvuuden monimuuttujaisiin analyysihin, joissa useat aikasarjat ovat toisiinsa sidoksissa, kuten energiankulutuksen ja sääolosuhteiden välisessä suhteessa.

Viimeaikaisissa tutkimuksissa SARIMA-malleja on edelleen kehitetty yhdistämällä ne koneoppimisen ja syväoppimisen menetelmiin. Esimerkiksi hybridimallit, joissa SARIMA yhdistetään neuroverkkoihin, voivat parantaa pitkän aikavälin ennustetarkkuutta ja tunnistaa monimutkaisempia riippuvuuksia. Näitä lähestymistapoja on sovellettu muun muassa sähköverkon kuormituksen ja taloudellisten syklien analysointiin.

### **2.2.3 Prophet**

Prophet on Facebookin kehittämä avoimen lähdekoodin työkalu aikasarjaennustamiseen, joka on suunniteltu erityisesti käsittelemään vahvasti kausiluonteisia aikasarjoja, joissa on useita kausia historiallista dataa. Se perustuu additiiviseen malliin, jossa epälineaariset trendit yhdistyvät vuosittaiseen, viikoittaiseen ja päivittäiseen kausivaihteluun sekä lomaefekteihin. Prophet on saatavilla sekä R- että Python-ohjelmointikielille, ja se tarjoaa nopean ja täysin automatisoidun ennustamisprosessin, jota käyttäjät voivat tarvittaessa hienosäätää.

Prophetin vahvuus on sen kyky käsitellä puuttuvia havaintoja ja äkillisiä trendimuutoksia sekä säilyttää luotettavuus poikkeamienkin keskellä. Se soveltuu erityisen hyvin aikasarjoihin, joissa esiintyy selkeää kausivaihtelua ja pitkää historiallista dataa. Työkalulla voidaan tuottaa nopeasti automatisoituja ennusteita, joita käyttäjä voi tarvittaessa jalostaa omiin tarpeisiinsa.

Yksi Prophetin merkittävistä ominaisuuksista on mahdollisuus huomioida lomien ja muiden erityistapahtumien vaikutukset ennusteissa. Tämä parantaa ennustetarkkuutta erityisesti liiketoimintaympäristöissä, joissa tällaiset tekijät voivat vaikuttaa merkittävästi kysyntään tai muihin mitattaviin suureisiin.

Prophetin avoimen lähdekoodin luonne ja sen saatavuus sekä R- että Python-kielille tekevät siitä helposti lähestyttävän ja laajalle käyttäjäkunnalle soveltuvan työkalun. Sen

joustavuus ja kyky käsitellä monimutkaisia aikasarjoja ovat tehneet siitä suosittua valinnan monilla aloilla, kuten liiketoiminnan suunnittelussa, talousennusteissa ja sosiaalisen median analytiikassa (Taylor & Letham, 2018; Taylor & Letham, n.d.).

## **2.3 Tekoäly ja koneoppiminen aikasarjaennustamisessa**

### **2.3.1 Transformer-arkkitehtuuri**

Transformer-arkkitehtuuri esiteltiin vuonna 2017 artikkelissa "Attention Is All You Need" (Vaswani ja muut, 2017). Tämä arkkitehtuuri on mullistanut aikasarjaennustamisen tarjoamalla tehokkaan tavan käsitellä sekvenssidataa ilman perinteisten toistettujen neuroverkkojen rajoituksia. Transformer perustuu täysin itsepäätteisyyteen (self-attention), mikä mahdollistaa rinnakkaisen laskennan ja pitkän aikavälin riippuvuuksien tehokkaan mallintamisen. Perinteiset mallit, kuten LSTM ja GRU, käsittelevät dataa sekventiaalisesti, mikä voi aiheuttaa laskennallisia pullonkauloja ja rajoittaa pitkän aikavälin riippuvuuksien havaitsemista. Transformerit sen sijaan hyödyntävät itsepäätteisyyttä, joka ottaa huomioon koko sekvenssin globaalin kontekstin jokaisessa laskentavaiheessa. Tämä parantaa sekä ennustetarkkuutta että laskennallista tehokkuutta.

Aikasarjaennustamisessa transformer-arkkitehtuurin soveltaminen on mahdollistanut merkittäviä edistysaskeleita. Esimerkiksi Informer-malli on suunniteltu käsittelemään pitkiä sekvenssejä tehokkaasti vähentämällä laskennallista monimutkaisuutta ja parantamalla pitkän aikavälin riippuvuuksien havaitsemista (Zhou ja muut, 2021). Informer hyödyntää itsepäätteisyyttä, joka keskittyy olennaisiin ajallisiin ominaisuuksiin ja ohittaa vähemmän merkitykselliset. Tämä lähestymistapa vähentää laskennallista kuormitusta ja parantaa ennustetarkkuutta, mikä tekee mallista erityisen soveltuvan pitkän aikavälin ennusteisiin suurissa ja monimutkaisissa datasarjoissa, kuten energian kulutuksen tai sääilmiöiden mallinnuksessa.

Autoformer on transformer-pohjainen aikasarjaennustusmalli, joka hyödyntää autokorrelaatiomekanismia yhdessä kausaalisen konvoluution (causal convolution) kanssa. Näiden kahden lähestymistavan yhdistäminen mahdollistaa sekä lyhyen että pitkän aikavälin riippuvuuksien entistä tarkemman tunnistamisen. Autokorrelaatiomekanismin avulla malli havaitsee tehokkaasti aikasarjojen toistuvat säännönmukaisuudet, kuten kausivaihtelut, kun taas kausaalinen konvoluutio osoittautuu erityisen hyödylliseksi lyhyen aikavälin piirteiden mallintamisessa. Näiden ominaisuuksien ansiosta Autoformer pystyy käsittelemään monimutkaisia ja laajoja aikasarjoja. Mallin soveltuvuus on jo osoitettu käytännössä esimerkiksi energiankulutuksen ja sääilmiöiden ennustamisessa (Wu ja muut, 2021). Vaikka transformer-arkkitehtuuri on tuonut merkittäviä parannuksia aikasarjaennustamiseen, sen soveltamisessa on edelleen haasteita. Laskennallinen monimutkaisuus ja suuri datan tarve voivat rajoittaa sen käyttöä erityisesti resurssirajoitteisissa ympäristöissä.

Transformer-arkkitehtuuri tarjoaa uuden tehokkaan lähestymistavan monimutkaisten ja pitkien sekvenssien käsittelyyn aikasarjaennustamisessa. Sen kyky mallintaa sekä lyhyen että pitkän aikavälin riippuvuuksia on parantanut ennustetarkkuutta monilla sovellusalueilla. Kuitenkin sen täysi hyödyntäminen edellyttää jatkotutkimusta ja -kehitystä, jotta laskennalliset vaatimukset voidaan optimoida ja sen potentiaali saadaan käyttöön erilaisissa ympäristöissä.

### **2.3.2 TimeGPT-malli**

TimeGPT on yksi ensimmäisistä yleiskäyttöisistä transformer-pohjaisista malleista, jotka kykenevät tuottamaan tarkkoja ennusteita ilman uudelleen koulutusta, jopa täysin uusilla datasarjoilla. Tämä zero-shot-päätely tekee siitä ainutlaatuisen työkalun monilla aloilla, sillä se ei vaadi raskaita mukautusvaiheita, kuten perinteiset menetelmät, esimerkiksi ARIMA ja LSTM, jotka usein edellyttävät sovelluskohtaisia säätöjä ja merkittävää koulutustyötä (Garza ja muut, 2024). TimeGPT:n arkkitehtuuri hyödyntää transformer-mekanismia yhdistäen pitkien ja monimutkaisten ajallisten riippuvuuksien tehokkaan mallintamisen laskennalliseen optimointiin. Lisäksi mallissa käytetty itsepääteisyys vähentää

huomattavasti perinteisten transformerien laskennallista kuormitusta (Zhou ja muut, 2021).

Transformer-pohjaiset mallit, kuten TimeGPT, ovat mullistaneet aikasarjaennustamisen erityisesti pitkän aikavälin riippuvuuksia sisältävissä sovelluksissa. Perinteiset mallit ovat usein haavoittuvia monimutkaisten kausaalisten suhteiden käsittelyssä, mutta transformerien kyky prosessoida suuria tietojoukkoja rinnakkain on mahdollistanut uudenlaisten lähestymistapojen kehittämisen. TimeGPT:n tehokkuus perustuu itsepäättelämekanismiin, joka keskittää huomion älykkäästi vain olennaisimpiin osiin datasta. Tämä parantaa ennustetarkkuutta ja laskennallista skaalautuvuutta, erityisesti suurikokoisissa ja monimutkaisissa datasarjoissa. Tämä tekee siitä ihanteellisen työkalun pitkän sekvenssin aikasarjaennusteissa, kuten sähkönkulutuksen ja sääilmiöiden mallinnuksessa (Garza ja muut, 2024; Zhou ja muut, 2021).

TimeGPT:n zero-shot-päätely erottaa sen merkittävästi muista malleista, kuten TimeLLM:stä, jotka edellyttävät lisäkoulutusta tai sovelluskohtaisia säätöjä (Paroha & Chotrani, 2024). Tämä ominaisuus tekee siitä erityisen arvokkaan ympäristöissä, joissa datan esikäsittely tai uusien mallien kehitys on haasteellista.

Mallin suorituskykyä on testattu laajasti eri sovellusalueilla. Energiateollisuudessa TimeGPT on parantanut sähköverkon kuormitusennusteiden tarkkuutta, mikä on vähentänyt ennustamisvirheitä ja mahdollistanut tehokkaamman resurssien suunnittelun. Tämä on johtanut merkittäviin kustannussäästöihin ja parempaan järjestelmästabiilisuuteen (Paroha & Chotrani, 2024). Öljy- ja kaasuteollisuudessa TimeGPT:tä on sovellettu sähköpumpppujen (ESP) kunnossapidon ennakointiin, jossa se on osoittautunut tarkaksi ja käytännölliseksi, tunnistuen kunnossapitotarpeet vähemmällä resursseilla kuin monet muut menetelmät.

Lisäksi TimeGPT on osoittanut potentiaalinsa ympäristöissä, joissa datan saatavuus on rajallista. Esimerkiksi sääilmiöiden pitkän aikavälin mallinnuksessa ja lääketieteellisissä

analyysitilanteissa malli on tarjonnut luotettavia ja helposti skaalautuvia ennusteita ilman tarvetta laajoille datakokonaisuuksille (Garza ja muut, 2024).

Vaikka TimeGPT:n suorituskyky on vakuuttavaa, sen käyttöönotossa on edelleen merkittäviä haasteita. Yksi keskeisistä rajoitteista on transformer-arkkitehtuurien korkea laskennallinen kuormitus, mikä rajoittaa mallin käyttöä pienissä organisaatioissa tai ympäristöissä, joissa laskentateho on rajallinen. Käytännössä tämä tarkoittaa, että monet käyttäjät voivat hyödyntää TimeGPT:tä vain API-rajapinnan kautta.

TimeGPT:n jatkokehityksessä pyritään erityisesti parantamaan mallin laskennallista tehokkuutta ja sovittamaan se ympäristöihin, joissa käytettävissä oleva datan määrä tai resurssit ovat rajalliset. Yksi lupaava lähestymistapa on hybridiratkaisujen kehittäminen, joissa yhdistetään transformer-teknologiaa kevyempiin laskennallisiin malleihin. Tällaiset ratkaisut voivat säilyttää TimeGPT:n tarkkuuden ja skaalautuvuuden samalla, kun ne mahdollistavat sen käytön myös resurssirajoitteisissa ympäristöissä (Garza ja muut, 2024).

### **2.3.3 Koneoppimisen sovellukset aikasarjaennustamisessa**

Koneoppimisen menetelmät ovat mullistaneet aikasarjaennustamisen tarjoten tehokkaita ratkaisuja monimutkaisten ja ei-lineaaristen ajallisten riippuvuuksien mallintamiseen. Sovellusalueet kattavat talouden, energian, terveydenhuollon ja teollisuuden, joissa koneoppiminen parantaa päätöksenteon optimointia ja lisää ennustetarkkuutta.

Bontempi ja muut (2013) esittelevät koneoppimisen strategioita aikasarjojen käsittelyyn ja korostavat yksiaskelisten ja moniaskelisten ennustamismenetelmien merkitystä. Yksiaskelisten menetelmien etuna on niiden yksinkertaisuus ja selkeys koulutusvaiheessa, kun taas moniaskeliset menetelmät, kuten rekursiiviset ja suorat ennustamismenetelmät, tarjoavat paremmat valmiudet pidemmän aikavälin riippuvuuksien mallintamiseen. Lisäksi paikalliset oppimismenetelmät mukautuvat dynaamisesti datan ominaisuuksiin, mikä tekee niistä erityisen tehokkaita ei-stationaaristen aikasarjojen ennustamisessa.

Syväoppimismenetelmät ovat nousseet keskeisiksi työkaluiksi erityisesti rahoitusalan sovelluksissa. Sezer ja muut (2020) osoittavat, että syväoppimismallit, kuten konvoluutioneuroverkot (CNN) ja pitkät lyhyen aikavälin muistiverkot (LSTM), ylittävät perinteiset koneoppimismallit ennustetarkkuudessa. Näiden mallien kyky oppia monimutkaisia riippuvuuksia ja havaita kaavoja laajoista dataseiteistä on merkittävä edistysaskel erityisesti osakekurssien, indeksien ja valuuttakurssien ennustamisessa.

Viime vuosina transformer-pohjaiset mallit, kuten Informer, Autoformer ja TimeGPT, ovat nousseet tärkeään asemaan aikasarjaennustamisessa. Nämä mallit hyödyntävät itsepääätteisyyttä (self-attention) pitkän aikavälin riippuvuuksien tehokkaaseen mallintamiseen ja käsittelevät suuria datamääriä laskennallisesti optimoidulla tavalla, mikä vähentää laskennallista monimutkaisuutta ja parantaa ennustetarkkuutta erityisesti pitkissä sekvensseissä (Garza ja muut, 2024; Wu ja muut, 2021; Zhou ja muut, 2021).

Koneoppimista sovelletaan laajasti eri aloilla:

**Energiasektori:** Kuormitusennustemallit hyödyntävät koneoppimista energian kysynnän optimoinnissa, mikä parantaa sähköverkkojen vakautta ja resurssien hallintaa (Kamalov ja muut, 2024).

**Terveydenhuolto:** Koneoppimiseen perustuvat ennustemallit, kuten Mayo Clinic Early Warning Score (MC-EWS), parantavat potilaan tilan heikkenemisen ennustamisen tarkkuutta, mahdollistaen varhaisen reagoinnin poikkeamiin (Romero-Brufau ja muut, 2021)

**Teollisuus:** Ennustavat huoltomallit, kuten TimeGPT ja Time-LLM, optimoivat laitteiden ylläpitoa esimerkiksi öljy- ja kaasuteollisuudessa, vähentäen kustannuksia ja parantaen tehokkuutta (Paroha & Chotrani, 2024; Eldele ja muut, 2024).

Koneoppimismallien soveltamisessa on edelleen merkittäviä haasteita, kuten suuri datan tarve, laskennallinen monimutkaisuus ja rajallinen yleistämiskyky pienissä dataympäristöissä. Näiden rajoitteiden vuoksi mallien käyttö voi olla haastavaa erityisesti resursisrajoitteisissa ympäristöissä.

Tulevaisuudessa tutkimuksen painopisteenä ovat kevyemmät mallit, jotka yhdistävät syväoppimisen ja perinteisten mallien vahvuudet. Näillä lähestymistavoilla pyritään vähentämään mallien laskennallista kuormitusta ja parantamaan niiden soveltuvuutta laajalle joukolle sovelluksia. Lisäksi siirto- ja nollakoulutusmenetelmät tarjoavat lupaavia ratkaisuja erityisesti tilanteisiin, joissa datan määrä on rajallinen tai uusien mallien kehittäminen ei ole kustannustehokasta (Jin ja muut, 2023; Garza ja muut, 2024).

## **2.4 Ennustemallien vertailu ja arviointi**

Ennustemallien vertailu ja arviointi on olennainen osa aikasarja-analyysiä, sillä se mahdollistaa erilaisten mallien suorituskyvyn objektiivisen tarkastelun ja parhaan ennusteen valinnan. Vertailun onnistumiseksi on tärkeää käyttää asianmukaisia mittareita ja menetelmiä, jotka huomioivat ennusteiden tarkkuuden, mallien monimutkaisuuden ja niiden soveltuvuuden analysoitavaan dataan.

Ennen kuin tarkastellaan ennustemallien suorituskyvyn mittareita ja tilastollisia vertailumenetelmiä, on tärkeää hahmottaa, miten tutkimuksessa käsiteltävät neljä ennustemallia: ETS, SARIMA, Prophet ja TimeGPT, eroavat toisistaan ominaisuuksiensa, vahvuuksiensa ja heikkouksiensa suhteen. Taulukko 2 kokoaa yhteen näiden mallien keskeiset piirteet ja antaa kokonaiskuvan siitä, missä tilanteissa kukin malli voi olla soveltuvin.

**Taulukko 2.** Yhteenveto neljästä tutkitusta ennustemallista.

<b>Malli</b>	<b>Kuvaus lyhyesti</b>	<b>Vahvuudet</b>	<b>Heikkoudet</b>
<b>ETS</b> ( <i>Exponential Smoothing</i> )	Muodostaa ennusteen painottamalla viimeisimpiä havaintoja vahvemmin (trendin ja kausivaihtelun käsittely).	<ul style="list-style-type: none"> <li>- Helppo toteuttaa ja tulkita</li> <li>- Soveltuu hyvin tilanteisiin, joissa data sisältää trendiä ja kausivaihtelua</li> <li>- Sopeutuu nopeasti uusiin havaintoihin</li> </ul>	<ul style="list-style-type: none"> <li>- Ei välttämättä sovellu monimutkaisiin tai ei-stationaarisiin aikasarjoihin</li> <li>- Herkkä äkillisille muutoksille, jos parametrien valinta ei onnistu</li> <li>- Parametrien hienosäätö voi viedä aikaa</li> </ul>
<b>SARIMA</b> ( <i>Seasonal ARIMA</i> )	Laajentaa ARIMA-mallia huomioimalla kausikomponentin. Sopii selkeän kausirakenteen dataan (esim. kuukausittain).	<ul style="list-style-type: none"> <li>- Joustava: yhdistää AR, I ja MA -prosessit kausivaihteluun</li> <li>- Vakiintunut tilastollinen menetelmä</li> <li>- Hyvä monissa perinteisissä liiketoimintaennusteissa</li> </ul>	<ul style="list-style-type: none"> <li>- Parametrien valitseminen voi olla työlästä (p, d, q, P, D, Q)</li> <li>- Soveltuvuus heikentyy, jos data sisältää monimutkaisia tai muuttuvia kausisijaksoja</li> <li>- Saattaa ylisovittua, jos malli on liian monimutkainen</li> </ul>
<b>Prophet</b>	Facebookin kehittämä additiivinen malli, jossa on valmiina trendi- ja kausikomponentit sekä lomien vaikutusten käsittely.	<ul style="list-style-type: none"> <li>- Käyttöönotto helppoa (automaattinen parametrien säätö)</li> <li>- Hyvä malli selkeille kausirakenteille ja tilanteisiin, joissa mm. lomat vaikuttavat merkittävästi</li> </ul>	<ul style="list-style-type: none"> <li>- Ei aina suoriudu kovin monimutkaisista trendi-kausivaihtelukuvioista yhtä hyvin kuin uudenlaiset mallit</li> </ul>
<b>TimeGPT</b>	Transformer-pohjainen malli, joka kykenee zero-shot-ennusteisiin sekä hienosäätöön.	<ul style="list-style-type: none"> <li>- Soveltuu suuriin ja monimutkaisiin aineistoihin</li> <li>- Zero-shot-ominaisuus vähentää koulutuksen tarvetta</li> <li>- Hienosäädöllä voidaan parantaa tarkkuutta</li> </ul>	<ul style="list-style-type: none"> <li>- Vaatii ulkoisen API-palvelun (lisenssi/maksu)</li> </ul>

### 2.4.1 Ennustemallien suorituskyvyn mittarit

Taulukossa 3 yleisesti käytetyt ennustenmallien mittarit.

**Taulukko 3.** Ennustemallien mittarit.

Mittari	Selitys
<b>MAE</b>	Mean Absolute Error (MAE): Keskimääräinen absoluuttinen virhe mittaa ennusteiden ja toteutuneiden arvojen välistä keskimääräistä poikkeamaa.
<b>RMSE</b>	Root Mean Squared Error (RMSE): Keskivirheiden neliöjuuri korostaa suurempia virheitä, mikä tekee siitä herkemman poikkeaville arvoille.
<b>MAPE</b>	Mean Absolute Percentage Error (MAPE): Keskimääräinen absoluuttinen prosentuaalinen virhe ilmaisee ennusteen suhteellisen tarkkuuden prosentteina.

Hyndman ja Athanasopoulos (2018) korostavat, että ennustemallien arviointiin käytettävien mittareiden valinta riippuu sovelluskohteesta ja datan ominaisuuksista. Esimerkiksi MAPE (Mean Absolute Percentage Error) on hyödyllinen silloin, kun halutaan ymmärtää ennusteen suhteellista tarkkuutta, mutta se voi antaa harhaanjohtavia tuloksia, jos data on nolliä tai hyvin pieniä arvoja.

Mallien arvioinnissa on tärkeää huomioida niiden monimutkaisuus. Liian monimutkainen malli saattaa ylisovittaa dataa oppimalla satunnaisen kohinan piirteitä, mikä heikentää sen kykyä yleistää uusiin havaintoihin. Ylisovittamisen välttämiseksi käytetään usein mallin valintakriteereitä, kuten Akaiken informaatiokriteeriä (AIC) ja Bayesin informaatiokriteeriä (BIC). Nämä kriteerit rankaisevat mallin monimutkaisuudesta ja auttavat löytämään tasapainon sovituskvyn ja yksinkertaisuuden välillä.

AIC- ja BIC-kriteerien käyttö edellyttää pääsyä mallin todennäköisyysfunktioon sekä tietoa sen parametrien lukumäärästä. Koska TimeGPT on suljettu, transformer-pohjainen kielimalli, nämä tiedot eivät ole saatavilla sen tuottamien ennusteiden perusteella. Tämän vuoksi AIC- ja BIC-arvoja ei voida laskea samalla tavalla kuin esimerkiksi ARIMA- tai

muiden eksplisiittisesti määriteltyjen mallien kohdalla, joissa parametrit ja todennäköisyysfunktio ovat tiedossa. Tässä tutkimuksessa TimeGPT:n arviointi perustuukin ennusteiden ja toteutuneiden arvojen välisten virhemittareiden sekä tilastollisten testien käyttöön, jotka eivät edellytä mallin sisäisten parametrien tuntemista.

Ristiinvalidointi on vakiintunut menetelmä ennustemallien arvioinnissa, sillä se tarjoaa kattavamman kuvan mallin suorituskyvystä kuin pelkkä yksittäisen opetus- ja testijoukon vertailu. Menetelmän yleinen toteutustapa on jakaa data useisiin osiin siten, että kussakin vaiheessa eri osat toimivat vuorollaan koulutus- ja testiaineistona. Tämän avulla voidaan tunnistaa sekä ylisovittamista että arvioida mallin kykyä säilyttää ennustetarkkuutensa eri ajankohtina.

Tässä tutkimuksessa ristiinvalidointi toteutetaan rolling-origin-lähestymistavalla, jossa data jaetaan peräkkäisiin koulutus- ja testijaksoihin. Kussakin jaksossa ennuste tuotetaan esimerkiksi TimeGPT-mallilla, ja sitä verrataan toteutuneisiin arvoihin käyttäen esimerkiksi MAE-mittaria. Menetelmä on erityisen soveltuva pitkän aikavälin aikasarjojen analysointiin, sillä se mahdollistaa ennustemallin suorituskyvyn arvioinnin eri hienosäätöaskelmäärillä (esim. 5, 30, 50, 100). Mikäli virhemittarin arvo pienenee vain marginaalisesti askelmäärän kasvaessa, lisäkoulutus ei todennäköisesti ole resurssitehokasta.

Ristiinvalidointi voi myös paljastaa tilanteita, joissa lisähienosäätö tuo vain vähäistä parannusta ennustevirheisiin samalla, kun laskenta-aika ja ylisovittamisen riski kasvavat merkittävästi. Tällöin hienosäädön hyöty voi jäädä vähäiseksi. Mikäli virhemittari pysyy lähes muuttumattomana eri askelmäärillä, on todennäköistä, että malli on saavuttanut niin sanotun ”saturaatiopisteen”, jolloin lisähienosäätö ei enää paranna tuloksia. Jos virhemittari, kuten MAE, ei parane merkittävästi hienosäätöaskelmäärän kasvaessa, lisähienosäätöä ei kannata jatkaa. Pieni virhe-eron kaventuminen voi olla laskennallisesti kallista ja kasvattaa ylisovittamisen riskiä. Usein on hyödyllisempää pyrkiä parantamaan datan laatua tai lisätä malliin uusia ominaisuuksia (features) kuin kasvattaa hienosäätöaskelia yli suositellun määrän.

Ristiinvalidoinnin ja systemaattisen virhemittarien vertailun avulla voidaan löytää optimaalinen hienosäätötaso. Näin voidaan varmistaa, että valittu malli yleistyy organisaation käytännön tarpeiden mukaisesti ja tuottaa tarkkoja ennusteita myös reaali maailman käyttötilanteissa.

#### **2.4.2 Tilastolliset testit ennustemallien vertailussa**

Ennustemallien vertailussa tilastolliset testit ovat keskeisiä työkaluja, jotka mahdollistavat mallien ennustetarkkuuden objektiivisen arvioinnin ja keskinäisen vertailun. Näiden testien avulla voidaan selvittää, onko havaittu ero mallien ennustevirheissä tilastollisesti merkitsevä vai johtuuko se satunnaisvaihtelusta.

Yksi yleisimmin käytetyistä testeistä ennustemallien vertailussa on Diebold-Mariano-testi, jonka Diebold ja Mariano esittelivät vuonna 1995. Testi arvioi kahden ennustemallin suorituskyvyn eroa vertailemalla niiden ennustevirheiden keskiarvoja. Tämä testi on erityisen hyödyllinen tilanteissa, joissa virheissä esiintyy toistuvia kuvioita tai virheiden suuruus vaihtelee epätasaisesti eri ajankohtina (Diebold & Mariano, 1995).

Diebold-Mariano-testin alkuperäistä versiota on kuitenkin kritisoitu sen herkkyydestä hylätä nollahypoteesi pienillä otoksilla. Harvey, Leybourne ja Newbold (1997) ehdottivat testiin korjausta, joka parantaa sen luotettavuutta pienillä otoskoolla. Korjaus huomioi otoskoon ja ennustehorisontin vaikutukset testin jakaumaan, vähentäen virheellisten hylkäysten riskiä.

Tilastollisten testien käyttö ennustemallien vertailussa tarjoaa systemaattisen tavan arvioida mallien suorituskyyä. Ne auttavat määrittämään, onko havaittu ero mallien välillä merkittävä vai satunnaisen vaihtelun tulosta. Tämä on erityisen tärkeää päätöksenteossa, jossa ennusteiden luotettavuus on keskeistä. On kuitenkin huomioitava, että tilastollisten testien tulokset voivat riippua otoskoosta, ennustehorisontista ja käytetystä virhefunktiosta. Siksi on suositeltavaa käyttää useita testejä ja tarkastella tuloksia kriittisesti ennen johtopäätösten tekemistä.

## 2.5 Ennustajien kokemukset ja teknologian omaksuminen

TimeGPT-ennusteiden integroiminen Power BI -alustaan edustaa merkittävää askelta kohdeorganisaation ennusteprosessin kehityksessä. Aiemmin ennusteet laadittiin pääasiassa perinteisin menetelmin ja asiantuntija-arvioihin nojaten. Uuden tekoälypohjaisen ennustemallin tavoitteena on kuitenkin automatisoida prosessia ja tarjota ennustajille apua visuaalisesti selkeän työkalun avulla. Työkalun avulla on mahdollista syventyä yksityiskohtiin (drill down) sekä vertailla rinnakkain budjetin, toteutuman ja TimeGPT:n tuottamia ennusteita. Ennusteiden uskottavuus, ymmärrettävyys ja helppokäyttöisyys ovat olennaisia tekijöitä, sillä ne määrittävät pitkälti, kuinka hyvin organisaation käyttäjät ottavat uuden mallin vastaan.

Teknologian hyväksyntää voidaan arvioida viitekehysten, kuten Technology Acceptance Model (TAM) ja UTAUT, avulla. Näiden mukaan käyttäjien omaksumiseen vaikuttavat erityisesti koettu hyödyllisyys, käytön helppous ja teknologian yhteensopivuus organisaation tavoitteiden kanssa. Lisäksi mallin visuaaliset ominaisuudet, uskottavuus ja vuorovaikutteisuuden mahdollisuudet vaikuttavat käyttäjien suhtautumiseen. Helppokäyttöisyys on keskeisessä roolissa tekoälyn hyödyntämisessä päätöksenteon tukena, ja teknisen tuen sekä koulutuksen merkitys korostuu erityisesti luottamuksen rakentamisessa uutta teknologiaa kohtaan.

### 2.5.1 Kvalitatiivinen tutkimus teknologian hyväksymisestä

Teknologian hyväksymisen teoriaa (Technology Acceptance Model, TAM) pidetään yhtenä keskeisimmistä viitekehyksistä käyttäjien teknologiakäyttäytymisen selittämisessä (Davis, 1989). Mallin perustana ovat kaksi muuttujaa: koettu hyödyllisyys (Perceived Usefulness, PU) ja koettu helppokäyttöisyys (Perceived Ease of Use, PEOU), jotka vaikuttavat käyttäjän asenteisiin, käyttöaikomuksiin ja lopulta itse käyttöön. Koettu hyödyllisyys kuvaa käyttäjän käsitystä siitä, parantaako järjestelmän käyttö hänen työtehoaan ja tuloksiaan, kun taas koettu helppokäyttöisyys viittaa siihen, miten vaivattomaksi ja yksinkertaiseksi käyttäjä kokee järjestelmän käytön (Davis, 1989).

Mallia on myöhemmin laajennettu malleilla kuten TAM2 (Venkatesh & Davis, 2000) ja UTAUT (Unified Theory of Acceptance and Use of Technology; Venkatesh ja muut, 2003), joissa on otettu huomioon esimerkiksi sosiaaliset normit, teknologian työn relevanssi sekä sen yhteensopivuus organisaation tavoitteiden kanssa. Näillä tekijöillä on merkittävä vaikutus teknologian omaksumiseen erityisesti organisaatioympäristöissä, joissa päätöksenteon, resurssienhallinnan ja kilpailukyvyyn parantaminen ovat keskiössä.

Tässä tutkimuksessa TimeGPT-mallin hyväksyntää tutkitaan kvalitatiivisin menetelmin, kuten kyselyllä ja avoimilla kommenteilla, jotka tarjoavat syvällistä tietoa teknologian omaksumiseen vaikuttavista tekijöistä. Tekoäly voi toimia laadullisen analyysin tukena esimerkiksi tunnistamalla toistuvia teemoja ja keskeisiä havaintoja käyttäjien palautteista (Christou, 2023). Tämän lähestymistavan avulla on mahdollista kartoittaa, miten TimeGPT koetaan päivittäisenä ennustetyökaluna ja miten sen käyttöä voidaan optimoida osana organisaation toimintaa ja päätöksentekoprosesseja.

### **2.5.2 Ennustajien suhtautuminen tekoälypohjaisiin ratkaisuihin**

Tekoälypohjaiset ratkaisut, kuten TimeGPT-malli, ovat yleistyneet ennustamisessa, ja niiden omaksuminen riippuu useista käyttäjäkohtaisista ja organisatorisista tekijöistä. Koettu hyödyllisyys ja helppokäyttöisyys ovat keskeisiä teknologian hyväksyntään vaikuttavia tekijöitä (Davis, 1989; Venkatesh ja muut, 2003). Koettu hyödyllisyys viittaa siihen, miten käyttäjät arvioivat mallin parantavan työtehoaan ja ennustetarkkuuttaan sekä vähentävän manuaalista työtä, kun taas koettu helppokäyttöisyys liittyy siihen, kuinka vaivattomaksi ja selkeäksi teknologian käyttö koetaan. Näiden molempien tekijöiden ollessa positiivisia käyttäjät todennäköisemmin hyväksyvät työkalun osaksi ennustekäytäntöjään.

Hyväksyntää voivat kuitenkin haastaa muutkin tekijät kuin hyödyllisyys ja käytettävyys. Sundar (2008) MAIN-mallin mukaan teknologiavälitteisen viestinnän arviointiin vaikuttavat muun muassa modaliteetti (esim. tulosten visuaalinen esitystapa), toimijuus (teko-

ällyn nähdään "toimijana"), vuorovaikutteisuus (käyttäjän mahdollisuudet vaikuttaa mallin toimintaan) ja navigoitavuus (teknologian intuitiivisuus). Ennustajien tapauksessa näihin kuuluvat esimerkiksi TimeGPT:n kyky integroitua muihin työkaluihin, sen tulosten uskottavuus ja se, miten selkeästi mallin ehdotukset voidaan tulkita ja hyödyntää päätöksenteossa.

Spiliotis (2023) painottaa, että vaikka tekoäly voi merkittävästi parantaa ennustetarkkuutta, käyttäjien on myös opittava luottamaan uudentyyppisiin ja usein monimutkaisiin mallirakenteisiin. Tämä korostaa tarvetta tekniselle tuelle ja käyttäjäkoulutukselle, jotka voivat vahvistaa luottamusta ja varmistaa, että tekoälypohjaisten työkalujen täysi potentiaali saadaan käyttöön.

Lisäksi Ali, Warraich ja Butt (2024) ovat havainneet, että helppokäyttöisyyteen ja suorituskykyyn liittyvät odotukset ovat keskeisiä tekoällyn hyväksynnässä eri aloilla, kuten koulutuksessa. Näitä periaatteita voidaan soveltaa myös organisaatioiden ennustustoiminnassa, jossa tukitoimet, kuten koulutukset ja helppokäyttöisyyden varmistaminen, ovat kriittisiä työkalujen onnistuneelle integroinnille.

TimeGPT:n käyttöönoton onnistumiseksi olisi hyödyllistä painottaa sekä teknologian tuomia konkreettisia hyötyjä että käyttäjien valmiuksia ymmärtää ja hyödyntää tekoällyn toimintoja. Samalla on tärkeää tukea ennustajien kokemaa luottamusta järjestelmän uskottavuuteen ja sen tuottamiin tuloksiin.

Yhteenvetona voidaan todeta, että TimeGPT-mallin onnistunut integrointi edellyttää perusteellista ymmärrystä käyttäjien kokemuksista, jotka kietoutuvat hyödyllisyyden, helppokäyttöisyyden, luotettavuuden ja ennusteiden uskottavuuden ympärille. Tekoälyllä tuettu analyysi voi auttaa tunnistamaan näitä kriittisiä tekijöitä ja mahdollistaa entistä kohdennetummat kehitys- ja tukitoimenpiteet, joilla TimeGPT:stä voidaan luoda entistä tehokkaampi päätöksenteon väline.

## 2.6 Aikaisemmat tutkimukset

Aikasarjaennustaminen on tärkeää useilla eri aloilla, kuten taloudessa, energia-alalla, terveydenhuollossa ja teollisuudessa. Viime vuosien tutkimukset ovat erityisesti painotuneet generatiivisten tekoälymallien, kuten TimeGPT:n, soveltuvuuden arviointiin ja vertailuun perinteisiin ennustemalleihin nähden. Tässä osiossa tarkastellaan keskeisiä aikaisempia tutkimuksia, jotka liittyvät TimeGPT:n ja muiden suurten kielimallien (LLM) hyödyntämiseen aikasarjaennustamisessa eri konteksteissa.

### 2.6.1 TimeGPT:n soveltaminen eri aloilla

Deforce, Baesens ja Asensio (2024) tutkivat TimeGPT:n soveltuvuutta älykkään maatalouden tarpeisiin erityisesti maaperän kosteuspitoisuuden ennustamisessa. Heidän tutkimuksensa osoitti, että TimeGPT saavutti parhaan ennustustarkkuuden, kun mallin hienosäätö perustui ainoastaan historiallisiin kosteuspitoisuusarvoihin. Eksogeenisten muuttujien lisääminen heikensi mallin suorituskykyä, mikä viittaa siihen, että TimeGPT:n esikoulutusdata ei täysin vastaa maatalouden käytännön tarpeita. Tutkimus korostaa TimeGPT:n potentiaalia maataloussovelluksissa, erityisesti tilanteissa, joissa käytettävissä oleva datamäärä on rajallinen. Tämä tukee myös kestäväen kehityksen tavoitteita, kuten SDG-6 ja SDG-12 (Sustainable Development Goals) (Deforce, Baesens & Asensio, 2024).

Liao ja muut (2024) arvioivat TimeGPT:n soveltuvuutta energiasektorilla sähkön kuormituksen ennustamiseen. Heidän tutkimuksensa osoitti, että TimeGPT suoriutui erityisen hyvin lyhyen aikavälin ennusteissa (1–6 tuntia), tarjoten tarkempia tuloksia kuin perinteiset mallit, kuten LSTM ja XGBoost. Kuitenkin pitkien, yli 24 tunnin ennustehorisonttien osalta perinteiset mallit osoittautuivat TimeGPT:tä tehokkaammiksi. Tämä viittaa siihen, että vaikka TimeGPT tarjoaa huomattavaa etua dataköyhissä ympäristöissä, sen kyvykyys ennustaa pitkän aikavälin trendejä kaipaavaa kehitystyötä (Liao ja muut, 2024).

Song ja muut (2024) esittelivät TimelyGPT-mallin, joka on kehitetty erityisesti terveydenhuollon pitkän aikavälin ennustetarpeisiin. TimelyGPT hyödyntää extrapoloivaa positio-

upotusta (xPos) ja konvoluutiokerroksia, mikä mahdollistaa sekä jatkuvien että epäsäännöllisesti näytteistettyjen aikasarjojen analyysin. Tutkimuksessa havaittiin, että TimeLyGPT pystyi ennustamaan kehon lämpötilan muutoksia ja unen vaiheiden siirtymiä jopa 6000 aikapisteen horisontilla. Tämä osoittaa mallin poikkeuksellisen kyvyn tarkkoihin pitkän aikavälin ennusteisiin (Song, Lu, Xu, Buckeridge & Li, 2024).

Paroha ja Chotrani (2024) vertailivat TimeGPT:tä ja Time-LLM:ää sähkösuokelluspumpujen (ESP) ennakoivassa kunnossapidossa öljy- ja kaasuteollisuudessa. Tutkimus osoitti, että TimeGPT saavutti hieman paremman tarkkuuden (95,2 %) kuin Time-LLM (93,6 %), erityisesti kriittisten mittareiden, kuten värähtelyn ja paineen alenemisen, tunnistamisessa. Tulokset korostavat TimeGPT:n merkitystä teollisuuden luotettavassa ja datavetoisessa kunnossapidon suunnittelussa (Paroha & Chotrani, 2024).

### **2.6.2 Suuret kielimallit aikasarjaennustamisessa**

Gruver, Finzi, Qiu ja Wilson (2023) tutkivat suurten kielimallien, kuten GPT-3:n ja LLaMA-2:n, suorituskykyä aikasarjaennusteissa zero-shot-tilanteissa ilman mallin hienosäätöä. Tutkimuksessa kehitetty LLMTIME-malli ylitti perinteiset aikasarjamallit, kuten ARIMA:n ja N-BEATS:n, useilla testidatoilla. LLMTIME:n kyky käsitellä kausivaihteluita ja multimaalisia jakaumia tekstimuotoisen tokenoinnin avulla paransi ennustetarkkuutta erityisesti talouden ja ympäristönhallinnan sovelluksissa (Gruver, Finzi, Qiu & Wilson, 2023).

Merrill ja muut (2024) analysoivat suurten kielimallien, kuten GPT-4:n, kyvykkyyttä ymmärtää ja analysoida aikasarjoja zero-shot-tilanteissa. Heidän tutkimuksensa osoitti, että vaikka mallit pystyvät tuottamaan ennusteita, niiden kyky suorittaa loogista päättelyä on rajallinen. Tämä pätee erityisesti etiologiseen päättelyyn ja monimutkaisiin aikasarjoihin liittyviin kysymyksiin, mikä rajoittaa mallien soveltuvuutta monimutkaisiin liiketoimintaympäristöihin (Merrill, Tan, Gupta, Hartvigsen & Althoff, 2024).

Zhang ja muut (2024) esittelivät LLMForecaster-mallin, joka yhdistää strukturoimattoman tekstidatan numeerisiin aikasarjatietoihin parantaakseen kausittaisten tapahtumien ennustamista. Malli paransi merkittävästi sesonkiin liittyvien tuotteen kysyntäennusteiden tarkkuutta, mikä osoittaa suurten kielimallien potentiaalin vähittäiskaupan ja muiden alojen kysynnänhallinnassa (Zhang ja muut, 2024).

### **2.6.3 Hybridimallit**

LSTPrompt-tutkimuksessa Liu, Zhao, Wang, Kamarthi ja Prakash (2024) esittelivät uuden lähestymistavan aikasarjojen ennustamiseen, jossa ennustustehtävät jaettiin lyhyen ja pitkän aikavälin osatehtäviin. Tämä menetelmä paransi sekä ennustustarkkuutta että mukautuvuutta, erityisesti tilanteissa, joissa datamäärä oli rajallinen. Lisäksi TimeGPT:n käyttö tuulivoimaloiden SCADA-datan ennustamisessa osoitti, että transformer-pohjaiset mallit voivat tarjota merkittäviä etuja teollisuuden kunnossapidossa, erityisesti dataköyhissä ympäristöissä (Fan, Song, Feng, Song, Liu & Jiang, 2024).

Demir ja Karaođlan (2024) vertailivat LSTM-autoenkooderia ja TimeGPT:tä aikasarjojen anomaliatunnistuksessa. Tutkimuksessa havaittiin, että vaikka LSTM-autoenkooderi saavutti paremman suorituskyvyn useimmissa datasarjoissa, TimeGPT tarjosi nopeamman laskenta-ajan ja zero-shot-ominaisuuden, mikä tekee siitä erityisen hyödyllisen ympäristöissä, joissa nopea käyttöönotto on välttämätöntä. Tulokset korostavat mallivalinnan merkitystä sovellustarpeiden ja datan ominaisuuksien perusteella sekä viittaavat mahdollisuuksiin kehittää hybridimalleja, jotka yhdistävät molempien lähestymistapojen vahvuudet (Demir & Karaođlan, 2024).

### **2.6.4 Teknologian omaksuminen ja käyttäjäkokemukset**

Tutkimuksissa on myös tarkasteltu, miten ennustajat ja organisaatiot omaksuvat uusia tekoälypohjaisia ennustemalleja. TimeGPT:n ja vastaavien mallien käyttöliittymien sekä visuaalisten esitysten merkitys on korostunut, sillä selkeät ja intuitiiviset visualisoinnit

lisäävät käyttäjien ymmärrystä ja vahvistavat luottamusta ennusteisiin. Lisäksi koulutuksen ja teknisen tuen merkitys on keskeinen mallien onnistuneessa käyttöönotossa. Tämä edellyttää organisaatioilta investointeja paitsi teknologiaan, myös käyttäjien osaamisen kehittämiseen (Paroha & Chotrani, 2024; Merrill ja muut, 2024).

### **2.6.5 Yhteenveto aikaisemmista tutkimuksista**

Aikaisemmat tutkimukset osoittavat, että TimeGPT ja muut suuret kielimallit tarjoavat merkittäviä etuja erityisesti dataköyhissä ympäristöissä ja lyhyen aikavälin ennusteissa. Esimerkiksi maatalouden kosteuspitoisuuden ja sähkön kuormituksen ennustamisessa TimeGPT on saavuttanut korkean ennustustarkkuuden niukoilla datamäärillä. Tämä perustuu sen kykyyn hyödyntää esikoulutettua tietoa, mikä vähentää riippuvuutta laajasta ja laadukkaasta datasta.

Lyhyen aikavälin ennusteissa suuret kielimallit ovat osoittaneet vahvuutensa erityisesti tilanteissa, joissa tarvitaan nopeaa reagointia ja tarkkoja ennusteita. Tällaisissa ympäristöissä niiden suorituskyky voi merkittävästi parantaa operatiivista tehokkuutta ja resursien hallintaa. Sen sijaan pitkän aikavälin ennusteissa ja monimutkaisissa riippuvuussuhteissa perinteiset mallit tai hybridimallit, jotka yhdistävät tilastollisia ja koneoppimisen menetelmiä, voivat osoittautua paremmiksi. Näiden lähestymistapojen vahvuus piilee perinteisten mallien tarkkuudessa pitkän aikavälin trendien ennustamisessa sekä suurten kielimallien joustavuudessa ja skaalautuvuudessa.

TimeGPT:n tehokas hyödyntäminen edellyttää lisäksi panostuksia käyttäjäkokemukseen, koulutukseen ja tekniseen tukeen. Selkeät käyttöliittymät, interaktiiviset visualisoinnit ja koulutus, joka kattaa sekä tekniset perusteet että käytännön sovellukset, parantavat ennustajien ymmärrystä ja luottamusta mallin toimintaan. Tekninen tuki puolestaan mahdollistaa nopean ongelmanratkaisun ja varmistaa, että mallista saadaan maksimaalinen hyöty.

## 2.7 Tutkimusaukkojen tunnistaminen

Tekoälypohjaisia ennustemalleja on tutkittu laajasti eri aloilla, kuten taloudessa, energiasektorilla ja terveydenhuollossa, mutta niiden käyttö liikenteenohjauksen taloudellisessa ennustamisessa on toistaiseksi jäänyt vähäisen tutkimuksen alueeksi. Aiemmat tutkimukset keskittyvät usein joko teknisten suorituskykyvertailujen tekemiseen tai tarkastelevat tekoälypohjaisia ratkaisuja muissa konteksteissa, kuten kysynnän ennustamisessa tai teollisuuden prosessinohjauksessa. Erityisesti TimeGPT-mallin hyödyntäminen organisaation sisäisissä ennusteprosesseissa, kuten kustannusten ennustamisessa, ja mallin vaikutukset ennustajien käyttäjäkokemukseen ja teknologian hyväksyntään, ovat jääneet vähäiselle huomiolle akateemisessa kirjallisuudessa.

Taloudellisten kustannusten ennustaminen aikasarjamalleilla on monissa tutkimuksissa painottunut myynnin, kassavirran tai osakemarkkinoiden liikkeiden ennustamiseen. Sen sijaan kustannusrakenteiden mallintaminen liikenteenohjauksen kaltaisessa, monimutkaisia palveluita tuottavassa toimintaympäristössä on edelleen tutkimusaukko, joka kaipaa tarkempaa tarkastelua. Transformer-pohjaisten mallien, kuten TimeGPT:n, soveltuvuudesta kustannusennustamiseen on saatavilla vain niukasti vertaisarvioituja julkaisuja. Tämänhetkinen tutkimus ei tarjoa kattavaa kuvaa mallien yleistettävyydestä tai niiden mahdollisista rajoituksista erityisesti kustannusdatan ennustamisessa.

Merkittävä osa julkaisuista keskittyy mallien tekniseen suorituskykyyn, esimerkiksi virhemittareihin ja ennustehorisontteihin, ilman, että huomioitaisiin ennustajien käytännön tarpeita, työnkulkuja tai kokemuksia tekoälypohjaisten työkalujen käytöstä. Tämän takia organisaation sisäisen teknologian omaksumisen, ennustajien luottamuksen ja työkalun käyttöön liittyvät tekijät, joita tarkastellaan esimerkiksi TAM- ja UTAUT-malleissa, jäävät usein tutkimuksen ulkopuolelle.

Tutkimusaukko voidaan määritellä kaksi osaan: TimeGPT:n hyödyntämisestä taloudellisten kustannusten ennustamiseen ei ole riittävästi tutkimustietoa, sekä mallin käytön vaikutuksia ennustajien kokemukseen, työnkulkuun ja organisaation päätöksentekoon ei

ole tarkasteltu kattavasti. Tämä tutkimus pyrkii laajentamaan nykyistä tietopohjaa juuri näiden keskeisten kysymysten valossa, tarjoten uusia näkökulmia sekä teknisen suorituskyvyn arviointiin, että käyttäjäkokemuksen ja teknologian hyväksynnän merkityksen ymmärtämiseen.

### 3 Tutkimusmenetelmät ja empiirinen osuus

Tässä luvussa käsitellään tutkimuksen metodologista lähestymistapaa, aineiston keruuta, datan esikäsittelyä sekä ennustemallien toteutusta ja vertailua. Tutkimus yhdistää kvantitatiivisen ja kvalitatiivisen tutkimusotteen, mikä voidaan luokitella mixed methods -lähestymistavaksi. Kvantitatiivisen osion tavoitteena on arvioida TimeGPT-mallin ennustetarkkuutta verrattuna perinteisiin tilastollisiin ja koneoppimispohjaisiin menetelmiin sekä kohdeorganisaation omiin ennusteisiin. Kvalitatiivinen osio pyrkii puolestaan lisäämään ymmärrystä ennustajien kokemuksista, teknologian hyväksymisestä ja tekoälyratkaisujen roolista organisaation päätöksenteossa.

Tutkimusstrategia etenee siten, että mallivertailu aloitetaan pienellä, selkeästi rajatulla aineistolla, minkä jälkeen tarkastelua laajennetaan konsernitasoisiin kustannusennusteisiin. Mallien suorituskykyä arvioidaan useiden mittarien avulla (MAE, MSE, RMSE ja MAPE), ja mallien välisiä eroja testataan tilastollisesti Diebold-Mariano-testillä (Harvey-Leybourne-Newbold (HLN) -korjauksella). Kvalitatiivinen aineisto (ennustajille suunnatun kyselyn vastaukset) analysoidaan teemoittelemalla ja soveltaen teknologian hyväksyntämalleja (esim. TAM ja UTAUT).

Tutkimuksen tavoitteena on muodostaa kokonaiskuva siitä, miten TimeGPT-malli suoriutuu kustannusten ennustamisesta perinteisiin menetelmiin verrattuna ja miten uuden ennusteteknologian käyttöönotto koetaan. Lopullisena päämääränä on esittää suosituksia TimeGPT-mallin jatkokehitykseen ja hyödyntämiseen sekä tuottaa uutta tietoa tekoälypohjaisten ennustemallien soveltuvuudesta taloudelliseen ennustamiseen liikenteenohjauskontekstissa.

#### 3.1 Aineiston keruu

Tutkimusaineisto koostuu kahdentyyppisestä datasta: (1) kvantitatiivisesta kustannusdatan joukosta ja (2) kvalitatiivisesta kyselyaineistosta. Kvantitatiivinen data on kerätty koh-

deorganisaation kirjanpito- ja laskentajärjestelmistä, ja se sisältää kuukausitasolla tarkasteltavia tietoja ajanjaksolta 1/2020–10/2024. Tämä ajanjakso tarjoaa riittävän pitkän aikasarjan monivuotisten trendien, kausivaihteluiden ja satunnaisten tekijöiden analysointiin. Tutkimuksessa tarkastellaan useita kustannustilejä eri hierarkiatasoilla (yritys-, vastualue- ja tilitaso), jotta ennustemallien skaalautuvuutta voidaan arvioida yksityiskohdaisista osa-alueista koko konsernia kattaviin summatason ennusteisiin.

Ensimmäisessä vaiheessa, jossa ennustemallien tarkkuutta verrataan perinteisiin menetelmiin, analyysi kohdistetaan kolmeen eri tarkkuustasoon: (1) kaikki yhtiöt, alin tiliyhmä (57 aikasarjaa), (2) yksi yhtiö, kaikki tasot (1853 aikasarjaa) ja (3) kaikki yhtiöt, kaikki tasot (5954 aikasarjaa). Tämä lähestymistapa mahdollistaa ennustemallien suorituskyvyn arvioinnin sekä pienemmässä, fokusoituneessa aineistossa että laajemmassa, monimutkaisessa kokonaisuudessa.

Tutkimuksessa käytetään kvalitatiivisen aineiston keräämiseen sähköistä kyselylomaketta, joka suunnataan kohdeorganisaation ennustajille maaliskuun 2025 ennustekierroksen päätyttyä. Kyselyn tavoitteena on kartoittaa TimeGPT-ennustemallin koettua hyödyllisyyttä, helppokäyttöisyyttä ja luotettavuutta sekä arvioida sen vaikutuksia ennustajien työprosesseihin. Näitä tavoitteita tukevat myös teknologian hyväksymisteorioiden, kuten TAM ja UTAUT.

TAM (Technology Acceptance Model) korostaa kahta keskeistä tekijää teknologian omaksumisessa, eli koettua hyödyllisyyttä (Perceived Usefulness) ja koettua helppokäyttöisyyttä (Perceived Ease of Use), jotka vaikuttavat suoraan käyttäjän asenteeseen ja aikomukseen käyttää uutta teknologiaa.

UTAUT (Unified Theory of Acceptance and Use of Technology) yhdistää useita teknologian hyväksynnän malleja ja huomioi mm. odotetun suorituksen (Performance Expectancy), odotetun vaivannäön (Effort Expectancy), sosiaalisen vaikutuksen (Social Influence) ja helpottavat olosuhteet (Facilitating Conditions).

Näiden teorioiden näkökulmasta kysely tarjoaa mahdollisuuden selvittää, miten ennustajat kokevat TimeGPT:n käytännön hyödyt ja käytön vaivattomuuden, sekä millainen merkitys järjestelmän luotettavuudella ja työyhteisön tuella on työkalun hyväksymiselle.

Kyselylomake toteutetaan Microsoft Forms -ympäristössä. Se lähetetään kaikille niille noin 118 ennustajalle, jotka vastaavat yrityksen eri tasoista ennusteista (yritystaso, vastualueet, projektit). Ennustekierroksen jälkeen jokainen vastaaja saa sähköpostitse linkin, jonka kautta hän pääsee suoraan vastaamaan kyselyyn. Forms-työkalu tallentaa tulokset automaattisesti yhteen tietokantaan (Excel-muotoinen tiedosto), mikä helpottaa sekä tulosten kokoamista että analysointia.

Kyselylomakkeen avulla kerättävä aineisto tarjoaa tutkimukselle laajan pohjan arvioida, kuinka hyvin TimeGPT tukee organisaation ennusteprosessia ja päätöksentekoa. Tuloksia hyödynnetään ennustemallin jatkokehityksessä sekä ennustajien osaamisen ja järjestelmän käytettävyyden parantamisessa. Analyysi toteutetaan kyselylomakkeen vastausten perusteella myöhemmässä vaiheessa tutkimusta, ja sen avulla pyritään tunnistamaan keskeiset kehittämistarpeet ja suositukset mallin tehokkaampaan hyödyntämiseen organisaation eri tasoilla.

### **3.1.1 Kyselylomakkeen toteutus ja sisältö**

Kyselylomakkeen suunnittelussa on hyödynnetty teknologian hyväksynnän teorioita (TAM, UTAUT) siten, että vastaajilta kysytään muun muassa:

- Miten hyödylliseksi (Perceived Usefulness) he kokevat TimeGPT-mallia tehdyn ennusteen?
- Miten helppokäyttöiseksi (Perceived Ease of Use) he kokevat ennustejärjestelmän käyttöliittymän, Power BI -pohjaisen ennusteiden esittämisen?
- Miten luotettavana (osittain liittyen UTAUTin sosiaaliseen vaikutukseen ja mallin uskottavuuteen) he pitävät tekoälypohjaista ennustetta?
- Kuinka paljon järjestelmä on jo integroitu päivittäiseen työskentelyyn ja millaista tukea (Facilitating Conditions) he vielä tarvitsevat?

Lomake rakennettiin Microsoft Forms-alustalle, koska se mahdollistaa linkkijakelun sähköpostitse, reaaliaikaiset tulosityhteenvedot ja suoran datan viennin Exceliin jatkoanalyysiä varten. Kysymyksiä on noin kymmenen: viisi-portaisilla asteikoilla mitataan hyödyn, helppouden ja luotettavuuden tasoa, ja avoimilla vastauslaatikoilla kerätään syventävää palautetta kehitystarpeista ja kohtaamistaan haasteista.

Kyselylomakkeen avulla on tarkoitus tutkia erityisesti:

- Ennustajien kokemus TimeGPT:n tarkkuudesta (tarkkuus ja luotettavuus).
- Käyttökokemus (helppokäyttöisyys, visuaalinen esitystapa, suodattaminen, po-  
rautuminen).
- Järjestelmän hyödyt ja rajoitteet budjetoinnissa ja päätöksenteossa (vaikutus re-  
surssien suunnitteluun).
- Mahdolliset kehitysehdotukset jatkokehitykseen (esim. uudenlaiset simulaatio-  
työkalut).
- Teknologian hyväksyntä (TAM/UTAUT): tarkastellaan, missä määrin järjestelmä  
koetaan aidosti hyödylliseksi ja helposti omaksuttavaksi, ja mitkä tekijät (esim.  
sosiaalinen tuki, organisaation kulttuuri) edesauttavat tai hidastavat käyttöä.

Tiivis yhdistelmä numeerisia ja laadullisia kysymyksiä tuottaa aineiston, jonka perus-  
teella voidaan arvioida ennusteraportin sekä TimeGPT-ennustemallin toimivuutta ja  
kohdentaa jatkokehitys vastaajien tarpeiden mukaisesti. Varsinainen kysely löytyy liit-  
teestä 8.

### **3.2 Datan esikäsittely**

Datan esikäsittely on keskeinen vaihe aikasarjaennustamisen onnistumisessa. Tässä tut-  
kimuksessa lähtökohtana oli hyödyntää mahdollisimman raakaa dataa, jolloin kustan-  
nusdataa ei merkittävästi siivottu tai suodatettu. Ainoana poikkeuksena olivat konserni-  
palvelun sisäiset veloitustilit, jotka jätettiin ennustedatan ulkopuolelle. Muut tilit, mu-  
kaan lukien sellaiset, joilla ei enää tarkastelujakson lopussa ollut kirjauksia, säilytettiin

mallien käytettäväksi. Tämän lähestymistavan avulla pyrittiin arvioimaan mallien kykyä tunnistaa trendien päättyminen tai arvojen vakiintuminen nollassa.

Datan esikäsittely toteutettiin pääasiassa Microsoft Excel -ympäristössä, jossa kustannusdata muokattiin tarvittavaan formaattiin ja tallennettiin CSV-tiedostoiksi. Aineistosta muodostettiin aikasarjoja siten, että jokaiselle hierarkiatasolle (yritys, vastuualue, tili) luotiin oma aikasarjansa. Mikäli data sisälsi puuttuvia havaintoja, ne jätettiin pääosin nolliksi tai käsiteltiin niin, että malli pystyi itse sovittamaan datan trendiin ja vaihteluihin. Tämä lähestymistapa valittiin, koska edistyneet mallit, kuten TimeGPT, kykenevät usein oppimaan myös meluisista aineistoista, mikä lisää mallien käytännön sovellettavuutta.

Joidenkin aikasarjojen osalta datan tarkempi siivoaminen, esimerkiksi poikkeamien tai erikoistilanteiden poistaminen, olisi ollut mahdollista. Tällaiset toimenpiteet olisivat kuitenkin voineet vääristää mallien todellista suorituskykyä käytännön olosuhteissa. Koska tutkimuksen tavoitteena oli arvioida ennustemallien toimivuutta realistisessa ympäristössä, jossa datan puhdistamiseen käytettävät resurssit ovat rajalliset, esikäsittely pidettiin tarkoituksella vähäisenä ja käytännönläheisenä.

Esikäsittelyn lopputuloksena syntyneet datasetit tarjoavat kattavan kuvan kohdeorganisaation kulurakenteesta ja sen kehityksestä vuosina 2020–2024. Nämä aineistot mahdollistavat ennustemallien laaja-alaisen vertailun ja tarjoavat tärkeää tietoa mallien suorituskyvystä eri hierarkiatasoilla sekä tilanteissa, joissa data on epätäydellistä. Näiden aineistojen pohjalta voitiin edetä ennustemallien rakentamiseen, niiden vertailuun ja suorituskyvyn arviointiin.

Lopullinen ennuste, josta ennustajilta pyydettiin palautetta kyselylomakkeella, laadittiin vielä yhtä ulottuvuutta (projekti) tarkemmalla tasolla. Ennusteen pohjana käytettiin toteutunutta aineistoa ajalta 2020–2025 helmikuuhun asti, ja ennuste laadittiin seura-

valle 22 kuukaudelle, joulukuuhun 2026 saakka. Ennustekierroksen tavoitteena oli arvioida mallien kykyä ennakoida tulevaa kehitystä sekä kerätä ennustajien asiantuntevaa palautetta käytetyistä menetelmistä.

### 3.3 Ennustemallit ja niiden toteutus

Tässä tutkimuksessa ennustemallien valinta ja toteutus perustuvat sekä perinteisiin tilastollisiin menetelmiin että uusimpiin tekoälypohjaisiin ratkaisuihin, erityisesti TimeGPT-malliin. Tavoitteena on luoda monipuolinen vertailuasetelma, jossa kukin malli ennustaa kohdeorganisaation kustannusdataa eri hierarkiatasoilla (yritys-, vastuualue- ja tilitaso) sekä erilaisilla aineiston laajuuksilla (esimerkiksi yhden yhtiön tilit vs. koko konsernin aikasarjat). Tämä lähestymistapa mahdollistaa mallien suorituskyvyn arvioimisen sekä yksinkertaisissa että monimutkaisissa ennustetehtävissä ja erilaisten datamäärien puitteissa.

Perinteiset tilastolliset mallit:

#### **ETS** (Exponential Smoothing)

ETS-malli on erityisen soveltuva ennustamiseen silloin, kun aineistossa esiintyy trendiä ja/tai kausivaihtelua. Se antaa suuremman painoarvon viimeisimmille havainnoille ja mukautuu nopeasti muuttuviin olosuhteisiin. Tässä tutkimuksessa ETS-mallin ennusteet toteutettiin Pythonin *statsmodels*-kirjaston avulla. Ohjelma-koodi löytyy liitteestä 5.

#### **SARIMA** (Seasonal ARIMA)

SARIMA-malli laajentaa perinteistä ARIMA-lähestymistapaa huomioimalla kausivaihtelut, jolloin se soveltuu monimutkaisen, säännöllisesti toistuvan vaihtelun mallintamiseen. SARIMA-ennusteet laadittiin *pmdarima*-kirjaston *auto\_arima*-toiminnolla, joka tunnistaa optimaaliset parametrit automaattisesti. Mikäli auto-

maattinen parametrien valinta ei tuottanut tyydyttäviä tuloksia, ennusteet laadittiin manuaalisesti asetetuilla malliparametreilla. Ohjelmakoodi löytyy liitteestä 3. SARIMA ennuste.

### **Prophet**

Metan kehittämä Prophet-työkalu on suunniteltu erityisesti vahvasti kausiluonteisille aikasarjoille. Se tukee trendi- ja kausikomponenttien automatisoitua mallintamista sekä mahdollistaa lomien ja muiden erikoistapahtumien huomioimisen ennusteissa (Taylor & Letham, n.d.). Prophet-ennusteet ajettiin Google Colab -ympäristössä, sillä tarvittavien kirjastojen asentamisessa paikallisympäristöön esiintyi teknisiä haasteita. Ohjelmakoodi löytyy liitteestä 4. Prophet ennuste.

Tekoälypohjaiset mallit:

### **TimeGPT zero-shot**

TimeGPT on transformer-pohjainen kielimalli, joka on suunniteltu erityisesti aikasarjojen ennustamiseen. Zero-shot-lähestymistavassa mallia käytetään sellaisenaan ilman hienosäätöä (fine-tuning) käyttäjän omalla datalla. Tämä lähestymistapa mahdollistaa mallin kyvyn arvioinnin generalisoida uusiin aikasarjoihin pelkästään esikoulutetun tietämyksen avulla. Zero-shot-lähestymistapaa käytettiin tutkimuksessa peruslähtökohtana mallin suorituskyvyn arviointiin. Ohjelmakoodi löytyy liitteestä 6.

### **TimeGPT hienosäätö (fine-tuning)**

TimeGPT-mallin suorituskykyä pyrittiin parantamaan hienosäädöllä kohdeorganisaation omalla aineistolla. Hienosäätö mahdollistaa mallin mukauttamisen organisaation kustannusrakenteen erityispiirteisiin, mikä voi merkittävästi parantaa ennustetarkkuutta. Tutkimuksessa hienosäätöä testattiin eri määrillä hienosäätö-

askelia (esim. 5–110 askelta), jotta voitiin arvioida hienosäädön laajuuden vaikutusta ennustetarkkuuteen. Näin pyrittiin optimoimaan mallin suorituskyky ottaen samalla huomioon aineiston ominaisuudet ja prosessiin liittyvät resurssivaatimukset. Ohjelmakoodi löytyy liitteestä 7 TimeGPT hienosäädetty ennuste.

Ennusteet laadittiin Python-ohjelmointikielellä pääosin paikallisella koneella Spyder-käyttöliittymässä (Anaconda). Prophet-mallin ennusteet ajettiin Google Colab -ympäristössä, joka tarjosi mallin tarvitseman tuen. Käytetyt Python-kirjastot olivat Statsmodels ja pmdarima (SARIMA- ja ETS-mallien sovitus ja ennustaminen), Prophet (fbprophet) Prophet-mallin ennusteisiin sekä Nixtla-kirjasto ja NixtlaClient API TimeGPT-mallin yhteyksien hallintaan.

Data esikäsiteltiin Microsoft Excelillä ja tallennettiin CSV-muotoon. Aikasarjatiedot muunnettiin yhdeksi matriisiksi ennen mallien sovittamista. Ennustetulokset tallennettiin takaisin CSV-tiedostoiksi tai kopioitiin suoraan Spyderin käyttöliittymästä. Pivotointien avulla tulokset saatiin samaan formaattiin alkuperäisdatan kanssa.

TimeGPT-ennusteet toteutettiin Nixtlan API:n avulla käyttäen NixtlaClient-rajapintaa. Ennusteen parametrit, kuten ennustehorisontti (useimmiten 10 kuukautta) ja hienosäätöaskelten määrä, annettiin käyttäjän toimesta. Zero-shot-ennusteissa TimeGPT hyödynsi esikoulutettuja painoja ilman lisäoptimointia, kun taas hienosäätöprosessissa malli mukautettiin kohdeorganisaation datan erityispiirteisiin. Nixtlaverse (2024).

Hienosäädetyissä malleissa tutkittiin hienosäätöaskelmien (5–110 askelta) vaikutusta ennustetarkkuuteen. Tulokset osoittivat, että virhemittareissa (MAE, MAPE, RMSE) saavutettiin pieniä parannuksia hienosäätöaskelten kasvaessa, mutta lisähienosäätö ei tuottanut merkittäviä hyötyjä tietyn rajan (n. 30–50 askelta) jälkeen. Tämä havainto tukee suositusta kustannustehokkaan ja rajatun hienosäädön käyttämisestä, jotta laskenta-aika ja resurssien käyttö pysyvät hallittuina.

Tutkimuksessa käytetyt ennustemallit kattavat laajan kirjon perinteisistä tilastollisista menetelmistä (SARIMA, ETS, Prophet) tekoälypohjaisiin transformer-malleihin (TimeGPT). Kaikki mallit sovitettiin samaan aineistoon, jonka muokkaaminen pyrittiin minimoimaan ennusteiden realistisuuden säilyttämiseksi. Toteutuksen aikana kehitetyt koodit automatisoivat mallien sovittamisen, ennusteiden laskennan ja tulosten tallentamisen.

Esimerkkikoodit (ks. liitteet) havainnollistavat, miten SARIMA-, ETS-, Prophet- ja TimeGPT-mallit toteutettiin käytännössä. Koodit sisältävät myös tilastollisia testejä, kuten Diebold-Mariano (DM) ja Harvey-Leybourne-Newbold (HLN) -korjauksen, joiden avulla mallien suorituskykyeroja voitiin arvioida objektiivisesti. Ennusteet yhdistettiin alkupe-  
räiseen dataan ja tallennettiin CSV-muotoon jatkoanalyysiä varten. Lopuksi ennustetu-  
lokset vietiin Power BI -alustalle visuaalista tarkastelua ja sidosryhmien käyttöä varten.

Tämän kokonaisvaltaisen menetelmän ja teknisen toteutuksen avulla tutkimus rakennettiin läpinäkyväksi, monipuoliseksi ja toistettavaksi. Näin varmistettiin, että TimeGPT:n suorituskykyä voitiin verrata luotettavasti perinteisiin menetelmiin ja tuloksia arvioida suhteessa organisaation käytännön tarpeisiin ja ennustajien kokemuksiin.

### **3.4 Data-analyysi**

Tässä tutkimuksessa data-analyysin tavoitteena on arvioida TimeGPT-mallin ja perinteisten aikasarjaennustemallien (ETS, SARIMA ja Prophet) sekä organisaation omien ennusteiden suorituskykyä kohdeorganisaation kustannusten ennustamisessa. Analyysissä keskitytään selvittämään, miten tarkasti kukin ennustemalli suoriutuu eri laajuustasoilla sekä punnitsemaan mallien välisiä eroja tilastollisesti. Lisäksi tarkastellaan, miten ennustetulokset voidaan saattaa loppukäyttäjien hyödynnettäviksi selkeässä ja visuaalisesti ymmärrettävässä muodossa.

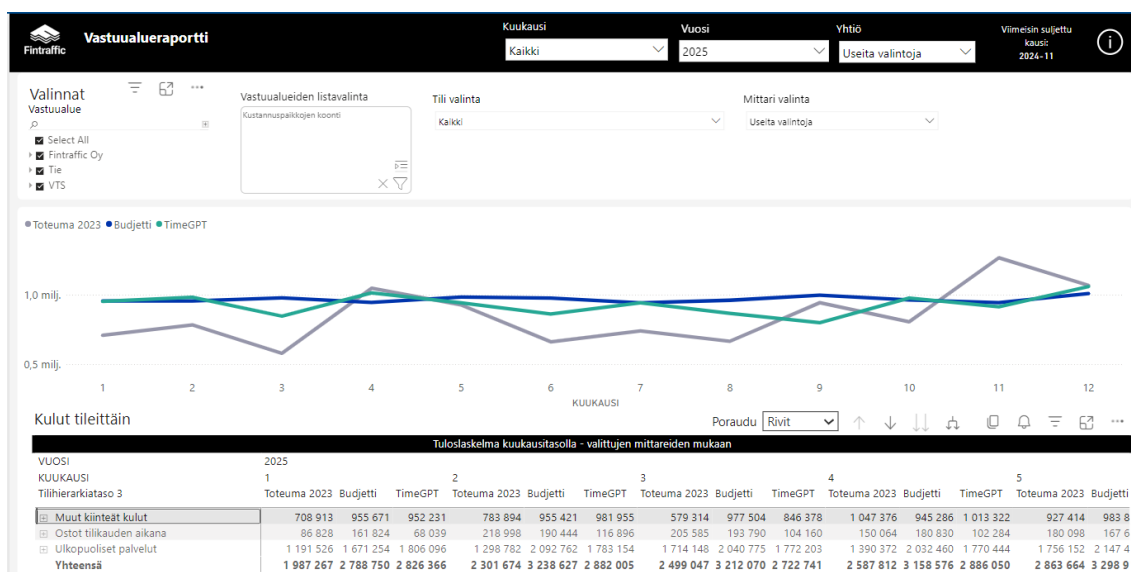
Tutkimuksen keskeisenä osana hyödynnetään useita ennustetarkkuuden mittareita. Katava joukko eritasoisia virhemittareita mahdollistaa mallien vertailun useasta näkökulmasta. Mittarit on valittu siten, että ne mittaavat sekä ennusteiden absoluuttista virhettä että suhteellisia poikkeamia. Lisäksi on huomioitu kuukausittaisten summien ja yksittäisten ennustearvojen tarkastelu eri mittausperiaatteilla. Käytetyt mittarit ovat: kk-MAE (kuukausittainen MAE), kk-MSE (kuukausittainen MSE), kk-RMSE (kuukausittainen RMSE) ja kk-MAPE (%) (kuukausittainen MAPE %).

Kuukausittaisiin (kk-) mittareihin viitataan, kun halutaan tarkastella mallin kykyä tuottaa tarkkoja ennusteita kuukausittaisista kokonaiskustannuksista. Kokonais-MAE puolestaan valottaa mallin suorituskykyä kaikilla ennustepisteillä erikseen, ilman yhteen niputtamista kuukausitasolle. Näiden erilaisten mittareiden avulla saadaan monipuolinen kuva mallien ennustetarkkuudesta.

Kokonais-MAE lasketaan siten, että jokaisen yksittäisen ennustetun arvon absoluuttinen virhe lasketaan suhteessa vastaavaan toteutuneeseen arvoon. Tämän jälkeen kaikki ennustetut arvot yhdistetään, virheiden summa jaetaan arvojen summalla, jolloin saadaan yleistävä käsitys mallin keskimääräisestä virheestä kaikilla ennustepisteillä. Kokonais-MAE ei perustu kuukausisummiin, vaan ottaa huomioon jokaisen ennustearvon erikseen, tarjoten näin yksityiskohtaisemman näkökulman mallien tarkkuuteen.

Ennustemallien vertailussa pelkkä virhemittarien tarkastelu ei aina riitä, sillä on mahdollista, että erot johtuvat satunnaisvaihtelusta. Siksi ennustevirheiden eroavaisuutta arviointiin tilastollisin testeillä, kuten Diebold-Mariano (DM) -testillä ja sen Harvey-Leybourne-Newbold (HLN) -korjauksella. Näiden testien avulla voidaan arvioida, ovatko havaitut erot mallien välillä tilastollisesti merkitseviä vai sattumanvaraisia. DM-HLN-testi mahdollistaa parivertailun eri mallien suorituskyvyn välillä, jolloin voidaan päätellä, onko esimerkiksi TimeGPT merkittävästi parempi tai heikompi kuin perinteiset mallit tietyissä olosuhteissa.

Ennusteet siirrettiin Power BI -alustalle, jossa ne esitettiin ennustajille Vastuualue-ra-  
portti-näkymän avulla (kuva 1). Työkalun yläosassa on interaktiivisia suodattimia yhtiölle,  
vastuualueille, tileille, kuukausille ja vuosille, mikä mahdollistaa joustavan tavan kohdis-  
taa tarkastelua. Pääkaaviossa esitetään kolme eri aikasarjaa: toteuma vuodelta 2023  
(harmaa viiva), budjetti (sininen viiva) sekä TimeGPT-ennuste (vihreä viiva). Näin käyttäjä  
havaitsee nopeasti poikkeamia toteuman, budjetin ja ennusteen välillä.



**Kuva 1.** Power BI ennustetyökalu

Kaavion alapuolella on taulukkonäkymä, josta käy ilmi vastaavat numerot kuukausittain,  
tilitasolla ja hierarkiatasolla. Käyttäjä voi porautua (drill down) kustannusten tarkempiin  
erittelyihin tai palata (drill up) yleisemmälle tasolle, mikä helpottaa erilaisten kustannus-  
komponenttien ja vastuualueiden vertailua. Työkalu tarjoaa myös suodatus- ja ryhmitte-  
lymahdollisuuksia, joiden avulla voi keskittyä esimerkiksi tiettyihin projekteihin tai kuu-  
kusiin. Samalla se tukee simulaatioiden ja vaihtoehtoisten ennustelaskelmien tekoa, jol-  
loin budjetointiin ja resurssisuunnitteluun liittyvät päätökset saadaan tehtyä ketteräm-  
min.

Ennustajat voivat tarvittaessa viedä Power BI:ssä näkyviä ennustetietoja talouden suunnittelutyökaluun, mikä parantaa ennusteiden hyödyntämistä budjetointi- ja resurssiprosesseissa. Lisäksi visuaalinen viiva- ja taulukkoyhdistelmä helpottaa erityisesti poikkeamien tunnistamista ja niiden juurisyiden etsimistä. Näin koko ennustekierroksen tiedot pysyvät samassa raportissa, tarjoten selkeän yleiskuvan mallien suorituskyvystä eri tasoilla.

Kvantitatiivisen analyysin rinnalla tutkimuksessa analysoitiin myös laadullista aineistoa, kuten ennustajille suunnattujen kyselyiden avoimia vastauksia. Laadullisen analyysin tavoitteena oli syventää ymmärrystä uuden teknologian käyttöönottoprosessista ja arvioida mallien tuottamien tulosten hyödyllisyyttä. Laadullinen aineisto täydensi kvantitatiivisia tuloksia ja auttoi tunnistamaan käyttäjäkokemukseen, teknologian hyväksymiseen ja tulosten tulkittavuuteen liittyviä teemoja.

TimeGPT:n suorituskyky analysoitiin eri tasoilla, ja mallien hienosäätöä testattiin vaiheittain. Hienosäätö osoitti erityisesti monimutkaisissa ja suurissa aineistokokonaisuuksissa parannuksia ennustetarkkuudessa. Tilastolliset testit osoittivat, että TimeGPT suoriutui tilastollisesti merkitsevällä tarkkuudella perinteisiin malleihin verrattuna, erityisesti suurissa ja kompleksisissa tietojoukoissa.

Tulosten visuaalinen havainnollistaminen ja käyttäjäkokemusten analysointi laadullisen aineiston avulla varmistivat, että löydökset eivät jääneet pelkästään numeeristen arvojen tasolle. Sen sijaan ne tukivat konkreettisesti mallien kehittämistä, käyttöönottoa ja hyödyntämistä osana organisaation strategista päätöksentekoa. Kokonaisvaltainen lähestymistapa tarjoaa käytännön työkaluja sekä jatkotutkimusaiheita, kuten hybridimallien kehittämistä ja ennustajien koulutuksen lisäämistä uusien teknologioiden tehokkaaseen hyödyntämiseen.

### 3.5 Tutkimusmenetelmiin ja dataan liittyvät eettiset näkökulmat

Tutkimuksen toteutuksessa on kiinnitetty erityistä huomiota eettisiin periaatteisiin ja tietosuojalainsäädännön noudattamiseen. Koska tutkimuksessa hyödynnetään organisaation sisäistä talousdataa ja henkilöstön kokemuksia, on tiedon luottamuksellinen käsittely, tarvittavien lupien hankinta ja osallistujien informoitu suostumus asetettu keskeiseksi lähtökohdiksi.

#### Tietosuoja ja luottamuksellisuus

- Kaikki tutkimuksessa käytettävä kustannusdata on käsitelty siten, ettei yksittäisiä henkilöitä tai heidän työpanostaan voida tunnistaa. Henkilötiedot ja muut yksilölliset tunnistetiedot on poistettu tai anonymisoitu ennen analyysiä.
- Anonymisoinnin yhteydessä aineistosta poistettiin sellaiset muuttujat, jotka voivat epäsuorasti paljastaa yksilöitä tai yksittäisiä toimijoita, kuten projektikohtaiset tai roolikohtaiset yksityiskohdat.
- Datan säilytys ja käsittely toteutettiin suojatuissa ympäristöissä, joihin on pääsy vain tutkimukseen osallistuvilla henkilöillä.

#### Luvat ja eettinen hyväksyntä

- Tutkimuksen aineiston käyttöön on haettu asianmukaiset luvat organisaatiolta (kohdeorganisaation johdolta). Tämä kattaa sekä talousdatan käytön että henkilöstölle suunnatun kyselyn toteutuksen.
- Tutkimuksen eettinen arviointi suoritettiin organisaation linjausten ja ohjeistuksen mukaisesti. Arvioinnissa varmistettiin, ettei tutkimusaineiston keruu tai analysointi aiheuta osallistujille yksityisyysriskejä tai muuta haittaa.

#### Ennustajien informoitu suostumus

- Kyselyyn osallistuvilta kohdeorganisaation ennustajilta varmistettiin informoitu suostumus. Osallistujille tiedotettiin tutkimuksen tarkoituksesta, aineiston käytöstä ja tietojen suojaustoimista ennen suostumuksen pyytämistä.

- Osallistuminen kyselyyn oli vapaaehtoista, eikä osallistumisesta tai osallistumatta jättämisestä aiheutunut taloudellisia tai muita seuraamuksia. Kyselyvastaukset käsiteltiin anonymisoituna siten, ettei yksittäisiä vastaajia voitu tunnistaa tuloksista.

#### **Kustannusdatan ja tekoälymallien käyttö**

- Tutkimuksessa hyödynnettiin tekoälypohjaista ennustusmallia (TimeGPT), joten datan käyttö suunniteltiin siten, ettei palveluntarjoaja (Nixtla) saanut pääsyä yksilöivään tai luottamukselliseen tietoon.
- Kaikki datan siirrot ja käsittelyt toteutettiin suojatulla yhteydellä (API-kutsut). Aineiston esikäsittelyssä varmistettiin, että mallille syötettiin vain yhteen niputettuja tietoja, jotka eivät sisällä tunnistettavia henkilöitä tai tapahtumia.

#### **Raportointi ja tulosten julkaisu**

- Tutkimuksen tulokset raportointiin siten, että yksittäisiä henkilöitä tai organisaation tarkkoja taloustietoja ei voitu tunnistaa. Käytetyt esimerkit ja visualisoinnit perustuvat anonymisoituun tai summatason dataan.
- Aineistoa ei jaettu kolmansille osapuolille, eikä sitä käytetty muihin tarkoituksiin kuin tässä tutkimuksessa määriteltyyn analyysiin.

Tutkimuksen toteutuksessa on huomioitu tietosuoja- ja tietoturvanäkökohdat sekä osallistujien oikeudet ja suostumus. Näin on varmistettu, että sekä talousdata että henkilöstön palautteet käsitellään eettisesti ja luottamuksellisesti, ja että tutkimus täyttää sekä akateemiset että lainsäädännölliset vaatimukset.

## 4 Tulokset

### 4.1 Ennustemallien suorituskyky

#### 4.1.1 Ennustemallien väliset tilastolliset erot

Tutkimuksessa vertailtiin eri ennustemallien tarkkuutta käyttäen Diebold-Mariano-testiä Harvey-Leybourne-Newbold -korjauksella. Testillä selvitettiin, ovatko erot ennustetarkkuuksissa tilastollisesti merkittäviä. Mitä pienempi testin p-arvo on, sitä varmemmin voidaan sanoa, että ennustemallien välillä on todellinen ero suorituskyvyssä. Tulokset esitetään taulukossa 4, jossa vertailtiin TimeGPT-malleja (zero-shot ja eri tavoin hienosäädetyt versiot), perinteisiä ennustemalleja (ETS ja SARIMA) sekä Prophet-mallia.

**Taulukko 4.** Diebold-Mariano-testin korjatut p-arvot.

Ennustemalli\Ennustemalli	TimeGPT zero-shot	TimeGPT hienosäätö 10 askelta	TimeGPT hienosäätö 30 askelta	TimeGPT hienosäätö 50 askelta	ETS	SARIMA	Prophet
TimeGPT zero-shot		0,663	0,959	0,535	0,041	0,005	0,132
TimeGPT hienosäätö 10 askelta	0,663		0,286	0,092	0,052	0,040	0,100
TimeGPT hienosäätö 30 askelta	0,959	0,286		0,023	0,109	0,149	0,101
TimeGPT hienosäätö 50 askelta	0,535	0,092	0,023		0,271	0,423	0,117
ETS	0,041	0,052	0,109	0,271		0,593	0,323
SARIMA	0,005	0,040	0,149	0,423	0,593		0,261
Prophet	0,132	0,100	0,101	0,117	0,323	0,261	

Tuloksista havaitaan, että TimeGPT zero-shot -mallin ennustetarkkuus poikkeaa merkittävästi perinteisistä menetelmistä, kuten SARIMA-mallista (p-arvo 0,0048) ja ETS-mallista (p-arvo 0,0408). Lisäksi havaittiin, että hienosäädettyjen TimeGPT-versioiden välillä (esim. 30 ja 50 hienosäätöaskelta) esiintyy eroa, mikä osoittaa mallin hienosäädön vaikuttavan ennustetarkkuuteen. Jo pieni hienosäätö (10 askelta) näyttäisi parantavan ennusteiden tarkkuutta esim. SARIMA-malliin verrattuna. Useimpien muiden mallien välillä erot eivät kuitenkaan olleet tilastollisesti merkittäviä (p-arvot yli 0,1). Tämä kertoo siitä, että monien mallien suorituskyky oli hyvin samankaltainen.

Yhteenvetona voidaan todeta, että p-arvojen analyysi paljastaa muutamia merkittäviä eroja TimeGPT:n ja perinteisten mallien sekä eri hienosäätöasteiden välillä, mutta monissa tapauksissa erot eivät olleet suuria. Siksi on tärkeää arvioida mallivalintaa myös käytännön näkökulmasta, kuten mallin ymmärrettävyyden, laskentatehon tarpeen ja sovellusympäristön näkökulmista. Diebold-Mariano-testin laskentaan käytetty koodi löytyy liitteessä 1.

#### **4.1.2 Ennusteet eri tasoilla ja ennustemalleilla**

Ennustemallien suorituskykyä eri aineistokokonaisuuksilla ja hierarkiatasoilla tarkasteltaessa havaittiin selkeitä eroavaisuuksia mallien kyvyssä tuottaa tarkkoja kustannusennusteita. Taulukoiden 5, 6 ja 7 tulokset osoittivat, että tilanteissa, joissa ennustettavia aikasarjoja oli suhteellisen vähän ja analysointi kohdistui yksinkertaisempiin tapauksiin (kaikki yhtiöt ja alin taso), perinteiset ennustemallit, kuten SARIMA, Prophet ja ETS, kykenivät tuottamaan TimeGPT-malliin verrattuna parempia tuloksia useilla mittareilla. Tämä näkyi erityisesti matalampina kk-MAE-, kk-MSE-, kk-RMSE- ja kk-MAPE-arvoina. Rajattu ja selkeä data sekä yksinkertaisemmat aikasarjarakenteet sopivat erityisen hyvin perinteisten menetelmien mallinnettaviksi.

Tilanteen muuttuessa monimutkaisemmaksi, esimerkiksi kun ennustettavien aikasarjojen määrä kasvoi huomattavasti (yksi yhtiö ja kaikki tasot sekä kaikki yhtiöt ja kaikki tasot), TimeGPT-mallin suorituskyky nousi selvästi esiin. Suuremmat aikasarjakokonaisuudet sisälsivät usein puutteellista dataa ja monimuotoisia piirteitä, kuten kuukausia ilman arvoja (kustannuksia) tai monimutkaisempia trendi- ja kausirakenteita. Tällaisissa olosuhteissa TimeGPT:n kyky käsitellä monimutkaisempia rakenteita, hyödyntää esikoulutettua tietoa ja hallita epätäydellistä dataa näkyi alempina virhemittareina. Esimerkiksi ”kaikki yhtiöt ja kaikki tasot” -tasolla TimeGPT saavutti merkittävästi tarkemmat tulokset useimilla mittareilla, erityisesti kk-MAE- ja kk-MAPE-arvoissa.

Myös TimeGPT:n hienosäätö (fine-tuning) paransi mallin suorituskykyä entisestään. Hienosäädettyjen mallien vertailu zero-shot-lähestymistapaan osoitti, että hienosäätö vähensi ennustevirheitä useissa tapauksissa. Tämä viittaa siihen, että TimeGPT:n mukauttaminen organisaation omaan kustannusdataan tuo selkeää lisäarvoa erityisesti monimutkaisissa ennustustilanteissa, joissa perinteiset mallit eivät enää tarjoa riittävää tarkkuutta.

Yhteenvetona voidaan todeta, että perinteiset menetelmät toimivat hyvin tilanteissa, joissa aineisto on yksinkertaista ja aikasarjojen määrä on rajallinen. Sen sijaan, kun ennustettavien aikasarjojen määrä kasvaa ja tilanne monimutkaistuu, TimeGPT-malli, erityisesti hienosäädettynä, osoittautuu merkittävästi paremmaksi vaihtoehdoksi. Tulokset vahvistavat käsitystä siitä, että uuden sukupolven tekoälypohjaiset transformer-mallit pystyvät hyödyntämään laajoja ja heterogeenisiä aikasarjakokonaisuuksia sekä käsittelemään puutteellista dataa tavalla, joka parantaa ennustetarkkuutta monimutkaisissa käytännön sovelluksissa.

**Taulukko 5.** Kaikki yhtiöt ja alin taso

Ennustemalli/Mittari	kk-MAE (t)	kk-MSE (mrd)	kk-RMSE (t)	kk-MAPE (%)	kokonais-MAE
TimeGPT zero-shot	645	504	710	8,67 %	<b>58 716</b>
TimeGPT hienosäätö 10 askelta	842	821	906	11,18 %	62 124
TimeGPT hienosäätö 30 askelta	727	628	792	9,67 %	61 061
TimeGPT hienosäätö 50 askelta	814	751	867	10,86 %	60 570
PROPHET	427	349	591	5,63 %	74 275
SARIMA	<b>248</b>	<b>135</b>	<b>368</b>	<b>3,45 %</b>	60 119
ETS	372	211	460	4,76 %	68 254

**Taulukko 6.** Yksi yhtiö ja kaikki tasot

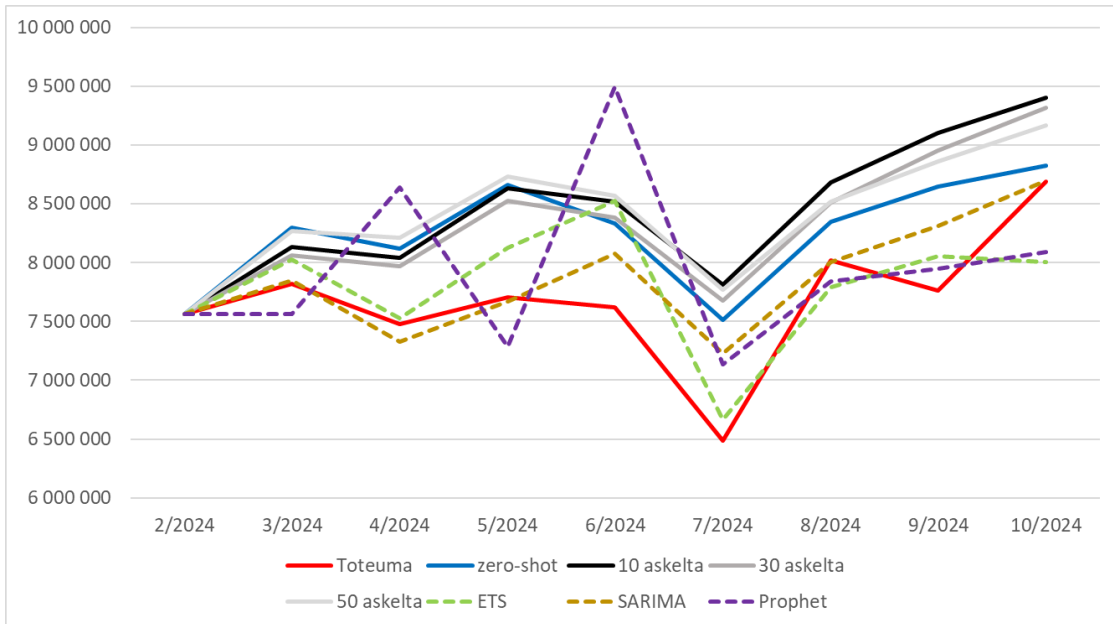
Ennustemalli/Mittari	kk-MAE (t)	kk-MSE (mrd)	kk-RMSE (t)	kk-MAPE (%)	kokonais-MAE
TimeGPT zero-shot	442	252	502	20,19 %	39 095
TimeGPT hienosäätö 10 askelta	309	134	366	13,78 %	37 199
TimeGPT hienosäätö 30 askelta	359	174	417	16,34 %	37 884
TimeGPT hienosäätö 50 askelta	<b>245</b>	<b>76</b>	<b>275</b>	<b>10,48 %</b>	<b>35 200</b>
PROPHET	435	270	520	20,38 %	64 713
SARIMA	874	823	907	38,46 %	46 736
ETS	710	565	751	31,27 %	63 354

**Taulukko 7.** Kaikki yhtiöt ja kaikki tasot

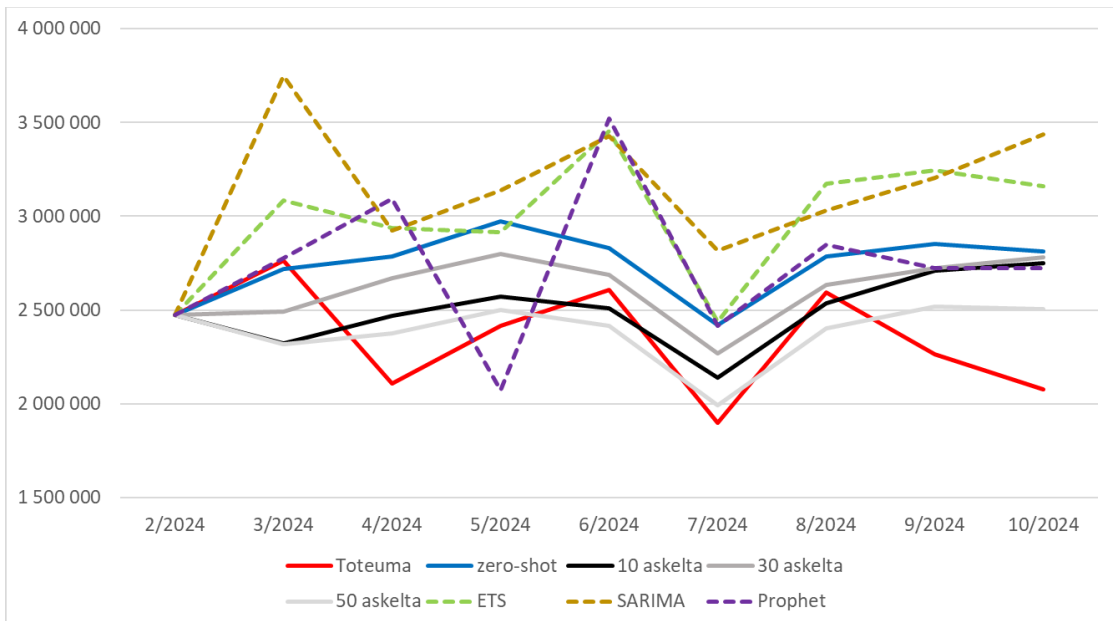
Ennustemalli/Mittari	kk-MAE (t)	kk-MSE (mrd)	kk-RMSE (t)	kk-MAPE (%)	kokonais-MAE
TimeGPT zero-shot	490	323	568	6,60 %	9 903
TimeGPT hienosäätö 10 askelta	<b>400</b>	<b>246</b>	<b>496</b>	<b>5,12 %</b>	9 542
TimeGPT hienosäätö 30 askelta	457	336	579	5,76 %	9 151
TimeGPT hienosäätö 50 askelta	604	524	724	7,57 %	<b>8 920</b>
PROPHET	667	565	751	8,55 %	15 913
SARIMA	875	883	939	11,70 %	11 009
ETS	962	1071	1035	12,70 %	14 988

Ennustekuvaajien (kuvat 2, 3 ja 4) tarkastelu tukee numeeristen mittarien perusteella tehtyjä havaintoja. ”Kaikki yhtiöt ja alin taso” -ennusteissa TimeGPT-mallin tuottamat ennustekäyrät sijoittuvat systemaattisesti toteumaa korkeammalle, mikä ilmenee yläkanttiin ennustamisena. Hienosäädettyjen TimeGPT-versioiden analyysi osoittaa kuitenkin selvästi, että hienosäätö parantaa ennusteen muotoa ja tarkkuutta suhteessa toteumaan. Samalla se säilyttää mallin kyvyn tunnistaa ja hahmottaa trendimuutoksia.

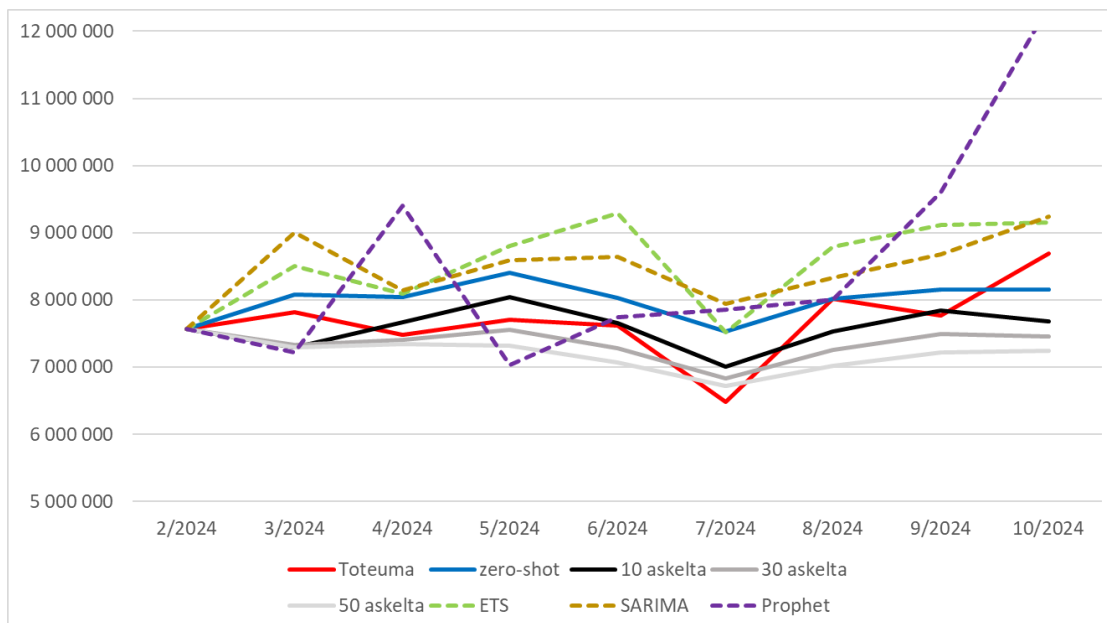
ETS- ja SARIMA-mallit onnistuvat myös tunnistamaan oleellisia kausittaisia ja trendinomaisia ilmiöitä, kuten esimerkiksi heinäkuulle ajoittuvan laskusuhdanteen. Prophet sen sijaan erottuu joukosta poikkeavalla ennustekäyrällään, eikä se kykene yhtä tarkasti mallintamaan trendien tai kausivaihteluiden käyttäytymistä. Tämä näkyy erityisesti ”Kaikki yhtiöt ja kaikki tasot” -ennusteen viimeisellä ennustejaksolla (10/2024), jossa Prophet tuottaa selvästi korkeampia kustannusarvioita verrattuna muihin malleihin ja toteumaan. Prophet-ennusteiden ajaminen useaan kertaan vahvisti tämän poikkeaman toistuvuuden, mikä korostaa mallin haasteita trendien ja kausivaihtelun luotettavassa mallintamisessa tässä ennusteympäristössä.



Kuva 2. Kaikki yhtiöt ja alin taso



Kuva 3. Yksi yhtiö ja kaikki tasot



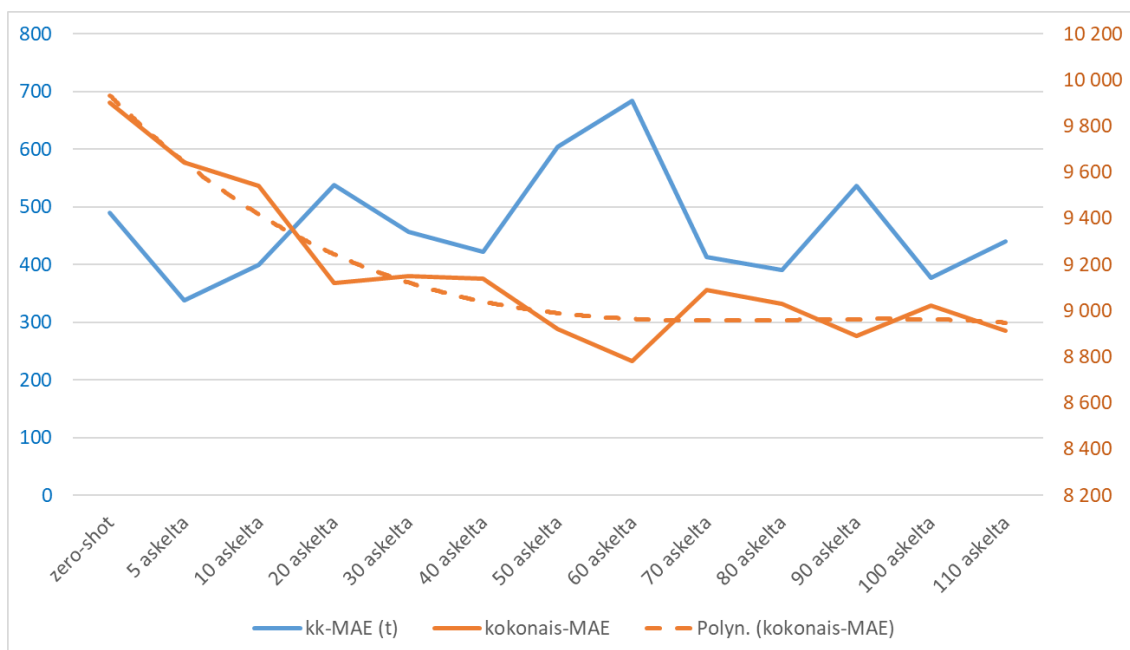
Kuva 4. Kaikki yhtiöt ja kaikki tasot

#### 4.1.3 TimeGPT:n hienosäädön askelmäärän vaikutus

Taulukon 8 arvoista sekä kuvasta 5 voidaan havaita, että kokonais-MAE laskee aluksi, kun hienosäädön askelmäärää kasvatetaan verrattuna zero-shot-lähtökohtaan. Erityisesti 5 askelta tuottaa merkittävästi matalamman kk-MAE:n ja kk-MSE:n kuin zero-shot, mikä viittaa siihen, että jo pienellä hienosäädöllä voidaan saavuttaa selkeä parannus ennustetarkkuuteen. Kokonais-MAE:n kehityksestä voidaan kuitenkin päätellä, että parannusta tapahtuu vain tiettyyn pisteeseen asti. Arvo laskee aluksi, mutta ei enää merkittävästi parane 60 askeleen jälkeen. Tämä viittaa siihen, että hienosäädön lisäys ei tuota loputtomasti tarkkuuden parannusta, vaan malli näyttää saavuttavan eräänlaisen saturaatiopisteen. Lisäksi kk-MAE pysyy melko tasaisella hyvällä tasolla noin 40 hienosäätöaskeleeseen saakka, minkä jälkeen ennusteen tarkkuus alkaa jälleen vaihdella. Tämä korostaa tarvetta tasapainoilla hienosäädön määrän ja laskennallisen tehokkuuden välillä, sillä tietyn rajan ylittäminen ei tuo merkittävää lisäarvoa tarkkuuden parantamiseen.

Taulukko 8. TimeGPT ennusteet eri hienosäädön askelmäärillä

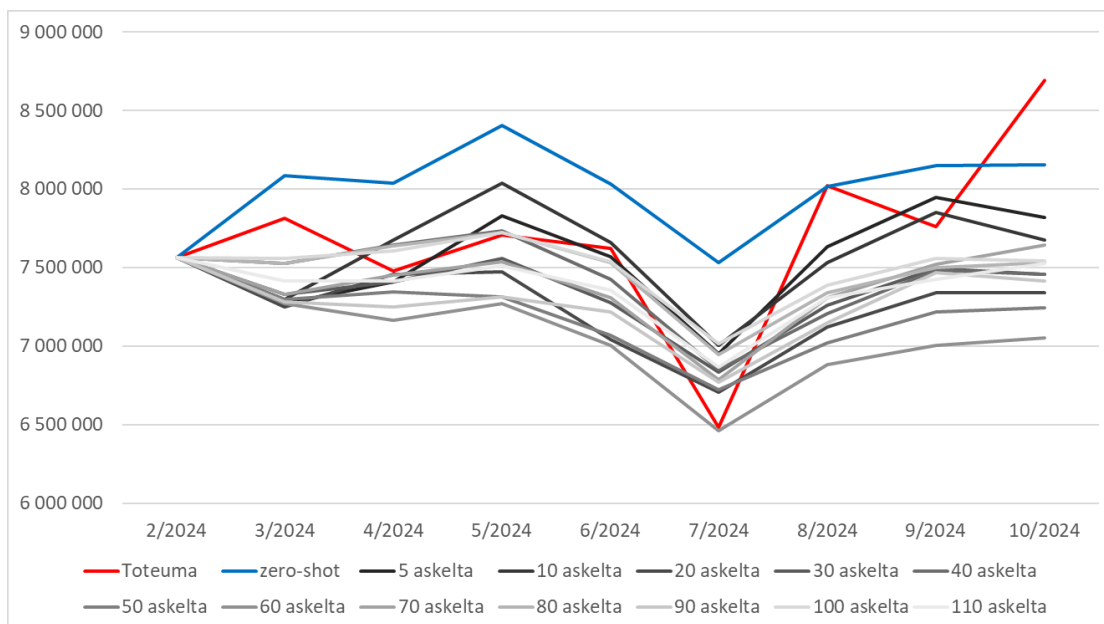
Ennustemalli/Mittari	kk-MAE (t)	kk-MSE (mrd)	kk-RMSE (t)	kk-MAPE (%)	kokonais-MAE
TimeGPT zero-shot	490	323	568	6,60 %	9 903
TimeGPT hienosäätö 5 askelta	<b>338</b>	<b>186</b>	<b>432</b>	<b>4,34 %</b>	9 642
TimeGPT hienosäätö 10 askelta	400	246	496	5,12 %	9 542
TimeGPT hienosäätö 20 askelta	538	447	669	6,73 %	9 120
TimeGPT hienosäätö 30 askelta	457	336	579	5,76 %	9 151
TimeGPT hienosäätö 40 askelta	422	320	565	5,31 %	9 140
TimeGPT hienosäätö 50 askelta	604	524	724	7,57 %	8 920
TimeGPT hienosäätö 60 askelta	684	690	831	8,51 %	<b>8 782</b>
TimeGPT hienosäätö 70 askelta	413	266	516	5,21 %	9 089
TimeGPT hienosäätö 80 askelta	391	276	525	4,95 %	9 031
TimeGPT hienosäätö 90 askelta	537	403	635	6,77 %	8 889
TimeGPT hienosäätö 100 askelta	377	268	518	4,79 %	9 020
TimeGPT hienosäätö 110 askelta	441	301	548	5,57 %	8 914



Kuva 5. TimeGPT ennusteiden kk-MAE ja kokonais-MAE eri hienosäädön askelmäärillä

Kuvan 6 tarkastelu tukee numeeristen arvojen perusteella tehtyjä havaintoja. Hienosäädön askelmäärän kasvaessa ennusteet näyttävät tasoittuvan, mikä voi viitata mallin kykyyn yleistää paremmin. Toisaalta tämä saattaa myös olla merkki ylisovittumisesta, jolloin malli menettää herkkyytensä tunnistaa yksittäisiä vaihteluita yhtä tehokkaasti kuin pienemmällä hienosäätömäärillä. Visuaalisesti ylisovittaminen ilmenee tasaisempuna ja

vähemmän reagoivana ennustekäyränä, joka ei välttämättä enää mukaile toteuman erityispiirteitä yhtä tarkasti. Tämä voi rajoittaa mallin soveltuvuutta tilanteissa, joissa lyhyen aikavälin vaihteluiden tarkka ennustaminen on kriittistä.



**Kuva 6.** TimeGPT ennusteet eri hienosäädön askelmäärillä

Tutkimuksessa hyödynnettiin myös aikasarjojen ristiinvalidointia (rolling-origin cross-validation) hyödyntäen TimeGPT-palvelua ja NixtlaClient-kirjastoa. Menetelmä pohjautuu Hyndmanin ja Athanasopoulosin (2018) suosittamaan käytäntöön, jossa data jaetaan useisiin koulutus- ja testijaksoihin "liukuvan ikkunan" periaatteella. Tämän lähestymistavan avulla arvioitiin mallin suorituskykyä eri ajankohtina ja tunnistettiin mahdollinen ylisovittaminen. Ohjelmakoodi: Ajallisen cross-validationin toteutus MAE-mittarilla, on esitetty liitteessä 2.

Ristiinvalidoinnin avulla saatuihin virhearvoihin peilattuna havaittiin, että hienosäädön askelmäärien lisääminen ei merkittävästi paranna mallin suorituskykyä. Virhemittari (MAE) laski vain marginaalisesti eri ennusteiden välillä, mikä viittaa käytännön kannalta melko pieneen parannukseen. Tämä havainto on linjassa kokonais-MAE:n ja muiden mittarien kanssa, jotka osoittavat, että hienosäädön lisäys tuottaa vain rajallisen hyödyn.

Eryityisesti yli 60–100 hienosäätöaskeleen tuoma lisäarvo osoittautui pieneksi, mikä tekee sen perustelun käytännön sovelluksissa haasteelliseksi. Malli näyttää vakautuvan jo varhaisessa hienosäätövaiheessa, eikä lisäpanostus hienosäätöön tuo merkittävästi tarkempia tuloksia. Tämä tukee johtopäätöstä, että loputon hienosäätö ei ole tarkoituksenmukaista. Sen sijaan huomio kannattaa kiinnittää muihin parannuskeinoihin, kuten datan rikastamiseen sekä siivoamiseen tai mallin rakenteellisiin muutoksiin, jotka voivat tarjota suuremman vaikutuksen ennustetarkkuuden parantamiseen.

#### **4.1.4 Ennustemallien ajoaikojen vertailu**

Ennustemallien ajoaikojen vertailu eri aineistokokonaisuuksissa (kaikki yhtiöt, alin tiliyhmä; yksi yhtiö, kaikki tasot; sekä kaikki yhtiöt, kaikki tasot) osoittaa merkittäviä eroja mallien laskennallisissa kustannuksissa ja käytännön soveltuvuudessa. Taulukon 9 perusteella TimeGPT-mallit, sekä zero-shot- että hienosäädetyt versiot (10–50 askelta), suoriutuvat ajallisesti erittäin tehokkaasti. Ennusteet valmistuivat kaikissa testatuissa aineistovariaatioissa vain noin 5–6 sekunnissa. Tämä osoittaa, että perinteiset mallit saattavat vaatia huomattavasti pidemmän laskenta-ajan laajoissa aineistoissa, kun taas TimeGPT:n nopeus tarjoaa käytännön sovelluksissa merkittävän edun erityisesti silloin, kun ennusteiden tuottaminen on toistuvaa ja aikaa rajoittavaa.

Perinteisten menetelmien kohdalla ajoajat venyivät merkittävästi datan monimutkaistuksessa. ETS-malli laski pienimmän aineiston kolmessa sekunnissa, mutta suurimmalla aineistolla laskenta-aika kasvoi 295 sekuntiin. SARIMA ja Prophet osoittautuivat vielä selvästi raskaammiksi: suurimmalla aineistolla SARIMA:lla kesti lähes 5 000 sekuntia ja Prophetilla yli 5 100 sekuntia. Huomionarvoista on, että Prophet-ennusteet toteutettiin Google Colab -ympäristössä, mikä saattoi vaikuttaa ajoaikoihin esimerkiksi käytettävissä olevan laskentatehon tai ajoalustan ominaisuuksien kautta.

On kuitenkin tärkeää huomioida, että TimeGPT on maksullinen palvelu, ja tutkimuksen tekohetkellä ilmainen käyttöoikeus kattoi vain rajallisen määrän ennusteajoja. Vaikka TimeGPT on teknisesti ja laskennallisesti kevyt vaihtoehto, sen laajamittainen hyödyntäminen saattaa edellyttää lisenssikustannuksia, mikä voi vaikuttaa sen kustannustehokkuuteen verrattuna paikallisesti ajettaviin perinteisiin menetelmiin. Jos ennustettavia aikasarjoja on kuitenkin runsaasti tai ennusteita on tuotettava usein ja nopeasti, TimeGPT:n nopeus ja helppo skaalautuvuus voivat tarjota merkittävän edun. Tämä edellyttää kuitenkin selkeää kustannus-hyötyanalyysiä ja perusteluja lisenssikustannuksille organisaation tarpeiden ja hyötyjen näkökulmasta.

**Taulukko 9.** Ennustemallien ajojen kesto (sekuntia)

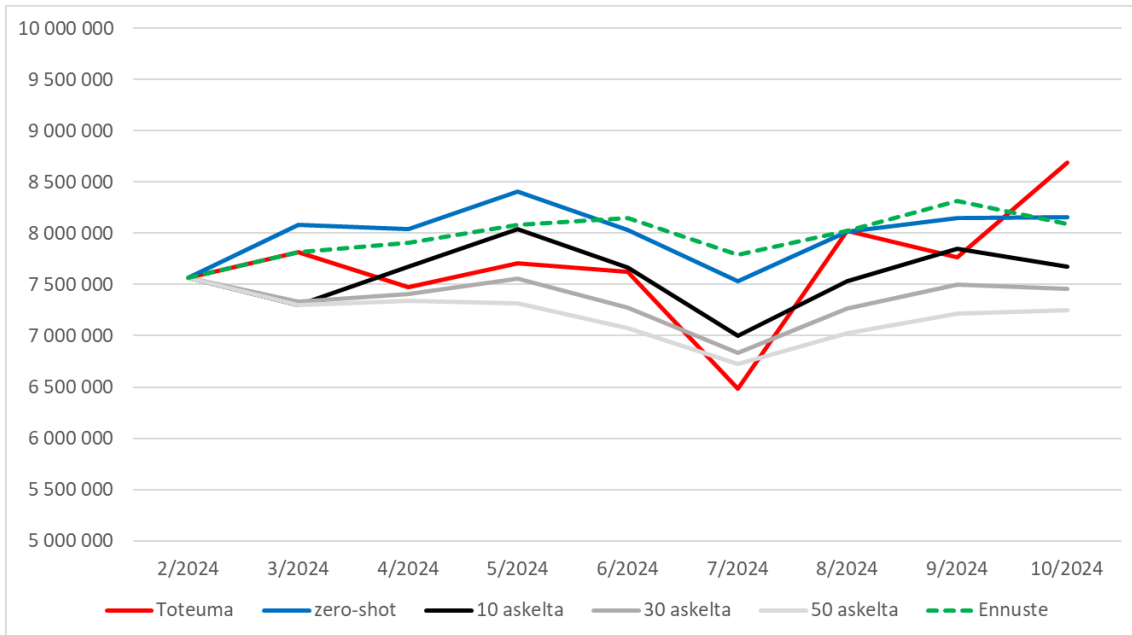
Ennustemalli\Aineisto	Kaikki yhtiöt, alin tiliryhmä	Yksi yhtiö, kaikki tasot	Kaikki yhtiöt, kaikki tasot
TimeGPT zero-shot	5	5	5
TimeGPT hienosäätö 10 askelta	5	5	5
TimeGPT hienosäätö 30 askelta	5	5	6
TimeGPT hienosäätö 50 askelta	5	5	6
ETS	3	84	295
SARIMA	48	1 544	4 962
Prophet	49	1 588	5 103

#### 4.1.5 Vertailu ennustajien aiempaan ennusteeseen

Edellä mainittujen TimeGPT-mallien ennusteita vertailtiin myös alkuvuodesta ennustajien laatimiin "K1-ennusteisiin" samoille kuukausille. Tulokset osoittavat, että ennustajien arviot (Ennuste) ovat usein hyvin lähellä TimeGPT:n zero-shot -ennusteita, mutta ne eivät ole yhtä tarkkoja kuin TimeGPT-mallin hienosäädetyt 10 askeleen tulokset. Esimerkiksi kk-MAPE-mittarin perusteella ennustajien tulos (6,48 %) on korkeampi kuin TimeGPT-mallin hienosäädöllä tuottama arvio (5,12 %). Tämä viittaa siihen, että hienosäätö on parantanut TimeGPT-mallin kykyä huomioida kustannusvaihteluita. Taulukko 10 ja kuva 7.

**Taulukko 10.** TimeGPT-ennusteet ja ennustajien tekemä ennuste

Ennustemalli/Mittari	kk-MAE (t)	kk-MSE (mrd)	kk-RMSE (t)	kk-MAPE (%)
TimeGPT zero-shot	490	323	568	6,60 %
TimeGPT hienosäätö 10 askelta	<b>400</b>	<b>246</b>	<b>496</b>	<b>5,12 %</b>
TimeGPT hienosäätö 30 askelta	457	336	579	5,76 %
TimeGPT hienosäätö 50 askelta	604	524	724	7,57 %
Ennuste	475	373	610	6,48 %

**Kuva 7.** TimeGPT-ennusteet ja ennustajien tekemä ennuste

On tärkeää huomioida, että tässä vertailussa käytetty ennustajien arvio oli vuoden ensimmäinen ja laadittu ilman tekoälymallin tukea. TimeGPT-ennusteet, jotka annettiin ennustajille ennustetyökalun avulla kvalitatiivisen tutkimuksen osana, kohdistuvat täsmälleen samaan ensimmäiseen ennustejaksoon. Näiden tulosten perusteella voidaan arvioida, että TimeGPT voi merkittävästi parantaa organisaation alkuvuoden ennustepohjaa.

## 4.2 Kvalitatiiviset havainnot

Tekoälypohjaisen TimeGPT-ennustemallin ja sitä hyödyntävän Power BI -raportin käyttöönoton arvioimiseksi toteutettiin 2025 vuoden ensimmäisen ennustekierroksen jälkeen sähköinen kysely kohdeorganisaation ennustajille. Kysely lähetettiin noin 118 ennustajalle, ja siihen vastasi 10 henkilöä (vastausprosentti n. 8,5%). Vastaajien roolit jakautuivat seuraavasti (Kysymys 1): 50% yritystason ennustajia, 30% vastuualuekohtaisia, 10% projektikohtaisia ja 10% konsernitason ennustaja/raportoija. Vaikka otoskoko oli pieni, vastaukset antavat arvokasta suuntaa antavaa tietoa käyttäjäkokemuksista ja teknologian hyväksynnästä. Tulokset löytyvät liitteestä 9. Kyselyn vastaukset.

Keskeinen havainto koski työkalujen käyttöastetta (Kysymys 2): Kymmenestä vastaajasta viisi (50%) ilmoitti hyödyntäneensä sekä Power BI -ennusteraporttia että TimeGPT-ennustetta. Neljä (40%) käytti ainoastaan raporttia, ja yksi (10%) ilmoitti käyttäneensä omia Excel-pohjiaan. Tämä tarkoittaa, että vain puolet vastaajista kokeili aktiivisesti TimeGPT-ennusteominaisuutta. Alhainen vastausprosentti ja työkalun osittainen käyttö voivat heijastaa sitä, että osa ennustajista ei kokenut työkalua vielä omakseen tai he tekivät ennusteensa perinteisin menetelmin ilman uutta työkalua, kuten tutkimuksen suunnittelu- vaiheessa arvioitiin mahdolliseksi. Analyysissä tämä jakautuminen otetaan huomioon.

### 4.2.1 Kokemusten analyysi

Analyysissä tarkastellaan erikseen Power BI -ennusteraportin ja TimeGPT-ennusteen käyttökokemuksia peilaten niitä teknologian hyväksyntämalleihin (TAM/UTAUT).

Power BI -raportin käyttökokemusta arvioitiin niiden yhdeksän vastaajan perusteella, jotka ilmoittivat käyttäneensä sitä (pois lukien Excel-käyttäjä). Kokonaisuutena Power BI -raportti sai erittäin positiivisen vastaanoton helppokäyttöisyytensä ja hyödyllisten ominaisuuksiensa ansiosta. Tämä on vahva signaali onnistuneesta PEOU:sta ja PU:sta TAM-mallin näkökulmasta.

- **Koettu helppokäyttöisyys** (PEOU, Kysymys 3): Raportin käyttö koettiin erittäin helppokäyttöiseksi. Asteikolla 1–5 keskiarvo kysymykselle "Kuinka helppokäyttöiseksi koit ennusteraportin käytön?" oli 4,0. Tämä korkea arvosana viittaa siihen, että raportin käyttöliittymä ja perustoiminnot olivat käyttäjille selkeitä ja intuitiivisia.
- **Hyödylliset ominaisuudet** (kysymys 4): Vastajat pitivät erityisen hyödyllisinä ominaisuuksina porautumismahdollisuutta eri hierarkiatasoille (mainittu 8/9 vastaajalla), graafista esitystapaa (6/9) ja suodatusvaihtoehtoja (4/9). Myös tiedonsiirto muihin työkaluihin mainittiin (3/9). Nämä ominaisuudet tukevat raportin käytettävyyttä ja tiedon analysointia.
- **Koettu hyödyllisyys** (PU) - Navita-ennusteen päivitys (Kysymys 7): Raportin hyödyllisyys Navita-ennusteen päivittämisessä sai kohtalaisen arvion, keskiarvon ollessa 3,44 (asteikolla 1-5) yhdeksän vastaajan kesken. Tämä viittaa siihen, että raportti tarjoaa tukea päivitysprosessiin, mutta ei välttämättä kata kaikkia tarpeita täydellisesti.

TimeGPT-ennusteeseen liittyviä kokemuksia arvioitiin pääasiassa niiden viiden vastaajan perusteella, jotka ilmoittivat käyttäneensä sitä:

- **Koettu tarkkuus ja luotettavuus** (kysymys 6): TimeGPT-ennusteen tarkkuus ja luotettavuus arvioitiin keskimäärin tasolle 3,6 (asteikolla 1-5) näiden viiden käyttäjän kesken. Arvio on lievästi positiivinen, mutta hajonta vastauksissa (arvot 3 ja 4) viittaa siihen, että kokemukset luotettavuudesta vaihtelivat tai olivat osin epävarmoja. Luottamus tekoälyn tuottamaan tietoon on keskeinen hyväksynnän tekijä.

- **Koettu hyödyllisyys (PU)** - Navita-ennusteen päivitys (kysymys 8): TimeGPT-ennusteen hyödyllisyys Navita-ennusteen päivittämisessä sai korkean arvion (keskiarvo 4,0/5,0) niiden viiden käyttäjän keskuudessa, jotka sitä käyttivät. Tämä eroaa merkittävästi kaikkien vastaajien keskiarvosta (3,44) ja osoittaa, että ne, jotka perehtyivät TimeGPT-ennusteeseen, kokivat sen selvästi hyödyllisemmäksi päivitystyössä kuin ne, jotka eivät sitä käyttäneet.
- **Vaikutus ennusteeseen ja suunnitteluun** (kysymys 9): Vaikutukset olivat kaksijakoisia. Osa käyttäjistä koki saavansa ajansäästöä (n. 20%), hyötyvänsä graafeista kuukausiallokaatioiden tarkastelussa tai pitävänsä ennustetta hyvänä kokonaisuuden tarkistuksena. Toisaalta yksi vastaaja, joka ei käyttänyt TimeGPT:tä, totesi sen logiikan olevan epäselvä, ja toinen (Excel-käyttäjä) totesi, ettei aikaa perehtymiseen ollut. Tämä korostaa, että koettu hyöty realisoituu vain, jos käyttäjä ymmärtää ja ottaa työkalun aktiiviseen käyttöön.

Avoimissa vastauksissa ja kehitysehdotuksissa nousi esiin useita tärkeitä teemoja:

- **Power BI -raportin kehitys** (kysymys 5): Toivottiin käyttöliittymän yksinkertaistamista ja räätälöintiä eri käyttötapauksiin. Lisäksi kaivattiin tarkempaa dataa ja läpinäkyvyyttä kirjanpidon tilien taakse (esim. kunnossapidon ja lämmityksen kustannuserien erittely), jotta ennustettavuus paranisi.
- **TimeGPT:n selittäminen ja validointi** (kysymys 5): Keskeinen toive oli TimeGPT-mallin logiikan avaaminen käyttäjille. Ehdotettiin, että mallin toimintaa ja tarkkuutta demonstroitaisiin vertaamalla sen tuottamia ennusteita aiempaan toteutumaan. Tämä lisäisi läpinäkyvyyttä ja auttaisi rakentamaan luottamusta. Myös delta-erojen selkeämpi visualisointi TimeGPT:n, budjetin ja toteuman välillä mainittiin.

- **TimeGPT:n toiminnalliset parannukset** (kysymys 5): Ehdotettiin ulkoisten muut-  
tujen huomioimista ennustetarkkuuden parantamiseksi ja mahdollisuutta siirtää  
TimeGPT-ennuste suoraan Navitaan napin painalluksella.
- **Koulutus ja tuki** (kysymys 10): Tarve ymmärtää TimeGPT:n logiikka korostui myös  
koulutustarpeissa. Kokemusten jakamista organisaatiossa pidettiin hyvänä tuki-  
muotona.
- **Jatkuvuus** (kysymys 11): Pidettiin tärkeänä, että tekoälymallia ylläpidetään ja ke-  
hitetään jatkossa.

Huolimatta esitetyistä kehitysehdotuksista ja haasteista, vastaajat olivat varsin haluk-  
kaita suosittelemaan työkalun hyödyntämistä kollegalleen. Kysymyksen "Suositteletko  
Power BI ennusteraportin ja TimeGPT-ennusteen hyödyntämistä työkaverillesi?" kes-  
kiarvo oli 4,0 (asteikolla 1-5) niiden yhdeksän vastaajan kesken, jotka käyttivät raporttia.  
Tämä korkea suosittelemahalukkuus viittaa siihen, että työkalussa nähdään potentiaalia ja  
sen tuomia hyötyjä arvostetaan.

Yhteenvetona laadullisesta analyysistä voidaan todeta, että Power BI -raportti on otettu  
hyvin vastaan ja sen helppokäyttöisyyttä arvostetaan. TimeGPT-ennusteominaisuus ja-  
kaa enemmän mielipiteitä: sitä aktiivisesti käyttäneet kokivat sen kohtalaisen luotetta-  
vaksi ja hyödylliseksi erityisesti Navita-päivityksissä, mutta merkittäväksi esteeksi nousee  
mallin logiikan epäselvyys ja tästä kumpuava luottamuspuola. Koulutuksen, selkeämmän  
dokumentaation ja kokemusten jakamisen tarve on ilmeinen.

#### 4.2.2 Ennustajien kokemukset

Kohdeorganisaation ennustajille suunnattu kysely tarjoaa, pienestä otoskoosta huoli-  
matta, arvokkaita näkemyksiä Power BI -pohjaisen ennusteraportin ja siihen integroidun  
TimeGPT-tekoälyennusteen vastaanotosta ja käytöstä organisaatiossa. Kokemukset ja-  
kautuivat selvästi raportin ja itse tekoälyominaisuuden välillä.

Power BI -raportti itsessään sai erittäin myönteisen vastaanoton ennustajilta. Sitä pidettiin erittäin helppokäyttöisenä (keskiarvo 4,0/5,0), mikä on keskeinen tekijä teknologian omaksumisessa (TAM: PEOU). Raportin arvoa analyysityössä tukivat hyödyllisiksi koetut ominaisuudet, kuten datan porautumismahdollisuudet, visuaalinen esitystapa ja suodatimet (TAM: PU). Lisäksi raportti koettiin kohtalaisen hyödylliseksi Navita-ennusteen päivittämisessä (keskiarvo 3,44/5,0), osoittaen sen integroituvan ainakin osittain olemassa oleviin työnkulkuihin.

TimeGPT-ennusteen kokemukset olivat vaihtelevampia ja perustuvat pääosin niihin viiteen vastaajaan, jotka aktiivisesti käyttivät ominaisuutta. Nämä käyttäjät arvioivat ennusteen luotettavuuden lievästi positiiviseksi (keskiarvo 3,6/5,0) ja kokivat sen erittäin hyödylliseksi Navita-ennusteen päivittämisessä (keskiarvo 4,0/5,0), mikä viittaa konkreettiseen apuun työssä (TAM: PU). Suurin kompastuskivi TimeGPT:n laajemmalle hyödyntämiselle oli kuitenkin sen toimintalogiikan epäselvyys, mikä herätti kysymyksiä luotettavuudesta ja ymmärrettävyydestä. Tämä haaste on suoraan yhteydessä koettuun helppokäyttöisyyteen (PEOU) ja hyödyllisyyteen (PU). Lisäksi vaikutukset työnkulkuihin vaihtelivat: osa koki ajansäästöä tai sai tukea analyysiin, kun taas ajanpuute tai ymmärryksen puute esti hyödyntämisen toisilla. Selkeä tarve lisäkoulutukselle, erityisesti mallin logiikan avaamiseksi, sekä kokemusten jakamiselle organisaatiossa nousi esiin, liittyen UTAUT-mallin "Facilitating Conditions" -tekijään.

Vastaajat antoivat konkreettisia kehitysehdotuksia sekä raportin käyttöliittymän ja datan esitystavan parantamiseksi että TimeGPT-mallin selittämiseksi, validoimiseksi ja integroimiseksi (kysymys 5). Huolimatta kritiikistä ja kehitystoiveista, yleinen halukkuus suosittelulla työkalua eteenpäin oli korkea (keskiarvo 4,0/5,0), mikä kertoo työkalussa nähdystä potentiaalista.

Yhteenvetona voidaan todeta, että ennustajien kokemukset osoittavat Power BI -raportin olevan onnistunut ja helppokäyttöinen työkalu. TimeGPT-tekoälymalli tarjoaa potentiaalia erityisesti niille, jotka perehtyvät sen käyttöön, mutta sen laajempi hyväksyntä

vaatii merkittäviä panostuksia läpinäkyvyyden lisäämiseen, mallin toiminnan selittämiseen ja käyttäjien koulutukseen ja tukemiseen. Teknologian käyttöönotossa teknisen suorituskyvyn rinnalla käyttäjäkokemuksen ja luottamuksen rakentamisen merkitys korostuu.

## 5 Johtopäätökset ja pohdinta

### 5.1 Analyysi

Tämä analyysi kokoaa tutkimuksen keskeiset havainnot TimeGPT-ennustemallin tehokkuudesta kohdeorganisaation kustannusennustamisessa verrattuna perinteisiin menetelmiin (ETS, SARIMA ja Prophet) sekä organisaation aiempiin käytäntöihin. Analyysissa integroidaan sekä kvantitatiiviset suorituskykymittaukset että kvalitatiiviset havainnot ennustajien käyttäjäkokemuksista.

#### 5.1.1 Suorituskyky suhteessa datan monimutkaisuuteen

Tutkimus osoitti, että optimaalinen ennustemalli riippui vahvasti käsiteltävän datan luonteesta. Kuten taulukoista 5–7 ilmeni, pienemmissä ja rakenteeltaan yksinkertaisemmissa aineistoissa perinteiset mallit, kuten SARIMA ja ETS, olivat usein TimeGPT:tä tarkempia mitattuna keskeisillä virhemittareilla. Diebold-Mariano-testin tulokset (Taulukko 4) kuitenkin osoittivat TimeGPT:n (zero-shot) eroavan tilastollisesti merkittävästi SARIMA- ja ETS-malleista.

TimeGPT:n edut korostuivat selvästi laajojen ja monimutkaisten aineistojen käsittelyssä (Taulukot 6 ja 7). Malli suoriutui perinteisiä menetelmiä paremmin tilanteissa, joissa data sisälsi puutteita, epäsäännöllisyyksiä ja monimutkaisia riippuvuuksia eri hierarkiatasoilla. Erityisesti laajimmassa aineistossa TimeGPT saavutti matalammat virheluvut (mm. kk-MAE, kk-MAPE), mikä vahvistaa sen kykyä hyödyntää esikoulutettua tietoa ja käsitellä heterogeenista dataa tehokkaasti. Prophet-malli osoittautui tässä tutkimuksessa muista malleista epävarmemmaksi, tuottaen ajoittain poikkeavia ennusteita.

### 5.1.2 Hienosäädön vaikutus ja optimaalinen taso

TimeGPT-mallin hienosäätö (fine-tuning) paransi ennustetarkkuutta systemaattisesti verrattuna zero-shot-lähestymistapaan. Jo vähäinen hienosäätö (5–10 askelta) tuotti merkittävän parannuksen (Taulukko 8). Tarkkuus parani hienosäätöaskelten määrää kasvatettaessa, mutta hyöty väheni noin 30–60 askeleen jälkeen, viitaten saturaatiopisteen saavuttamiseen. Myös ristiinvalidointi vahvisti havainnon, että lisäaskleet eivät tuottaneet merkittävää parannusta virhemittareihin. Liiallinen hienosäätö voi myös johtaa mallin ylisovittumiseen tai herkkyyden menettämiseen datan vaihteluille. Tulokset puoltavat rajatun ja kustannustehokkaan hienosäädön käyttöä.

### 5.1.3 Laskennallinen tehokkuus ja käytännön näkökohdat

Laskennallisessa tehokkuudessa TimeGPT oli ylivoimainen. Se tuotti ennusteet sekunneissa kaikilla aineistoilla, kun taas perinteiset mallit vaativat laajoilla aineistoilla minuutteja tai jopa yli tunnin (Taulukko 9). Tämä nopeus on merkittävä käytännön etu toistuvassa ja laajamittaisessa ennustamisessa. Huomioon otettava tekijä on kuitenkin TimeGPT:n mahdollinen API-pohjainen käyttökustannus verrattuna paikallisesti ajettaviin ilmaiisiin malleihin. Koska tilastolliset testit (Taulukko 4) osoittivat monien mallien suorituskyvyn olevan samankaltainen, käytännön tekijät kuten laskenta-aika ja käyttökustannukset nousevat tärkeään rooliin mallia valittaessa.

### 5.1.4 Vertailu aiempiin ennustekäytäntöihin

Verrattaessa TimeGPT-mallia ennustajien aiemmin ilman tekoälytukea laatimaan K1-ennusteeseen, hienosäädetty TimeGPT (10 askelta) osoittautui tarkemmaksi (Taulukko 10). Esimerkiksi kk-MAPE oli TimeGPT:llä 5,12 % verrattuna ennustajien 6,48 %. Tämä tulos viittaa siihen, että TimeGPT voi tarjota organisaatiolle luotettavamman ja tarkemman perustan ennusteprosessille, parantaen ennusteiden laatua verrattuna aiempiin manuaalisiin menetelmiin.

### 5.1.5 Käyttäjäkokemukset ja teknologian hyväksyntä

Kvalitatiivinen kysely osoitti, että ennustajat kokivat Power BI -pohjaisen ennusteraportin erittäin helppokäyttöiseksi (korkea PEOU) ja hyödylliseksi (korkea PU) sen toiminnallisuuksien, kuten visualisoinnin ja porautumisen, ansiosta. Sen sijaan siihen integroituun TimeGPT-tekoälyennusteeseen suhtauduttiin vaihtelevammin: vaikka ne harvat, jotka sitä aktiivisesti käyttivät, pitivät sitä kohtalaisen luotettavana ja hyödyllisenä erityisesti Navita-päivityksissä, sen toimintalogiikan epäselvyys ja läpinäkyvyyden puute nousivat merkittäväksi esteeksi laajemmalle hyväksynnälle, heikentäen luottamusta sekä koettua helppokäyttöisyyttä ja hyödyllisyyttä. Keskeiseksi havainnoksi nousi, että teknologian hyväksyntä riippuu vahvasti käyttäjien luottamuksesta ja ymmärryksestä, minkä vuoksi mallin logiikan selittämisen ja koulutuksen tarve korostui (UTAUT: Facilitating Conditions), vaikka työkalun potentiaali tunnistettiin korkeana suositteluhalukkuutena. Näin ollen Power BI on onnistunut alusta ennusteiden esittämiseen, mutta TimeGPT:n teknisen potentiaalın täysimääräinen hyödyntäminen organisaation arjessa vaatii merkittäviä panostuksia läpinäkyvyyden lisäämiseen ja luottamuksen rakentamiseen.

### 5.1.6 Yhteenveto analyysista

Analyysin perusteella TimeGPT on teknisesti tehokas ja laskennallisesti erittäin nopea työkalu kohdeorganisaation kustannusennustamiseen, erityisesti kun käsitellään suuria, monimutkaisia ja mahdollisesti epätäydellisiä aikasarja-aineistoja. Näissä tilanteissa se suoriutui kvantitatiivisesti perinteisiä malleja (ETS, SARIMA, Prophet) paremmin. Mallin hienosäätö paransi tarkkuutta, mutta optimaalisen hyödyn saavuttamiseksi riittää maltillinen hienosäätö (n. 30-50 askelta). Lisäksi TimeGPT osoitti potentiaalia parantaa ennusteiden tarkkuutta verrattuna organisaation aiempiin manuaalisiin ennustekäytäntöihin.

Kuitenkin pelkkä tekninen suorituskyky ei riitä. Kvalitatiiviset havainnot korostivat, että vaikka ennusteiden esittämiseen käytetty Power BI -raportti koettiin erittäin helppokäyt-

töiseksi ja hyödylliseksi, itse TimeGPT-tekoälymallin hyväksyntä kohtasi haasteita. Keskeisin este oli mallin toimintalogiikan epäselvyys, joka heikensi käyttäjien luottamusta ja halukkuutta hyödyntää ennustetta täysimääräisesti, vaikka aktiivikäyttäjät kokivatkin sen hyödylliseksi tietyissä tehtävissä.

Kokonaistulkintana voidaan todeta, että TimeGPT tarjoaa merkittävän teknisen edun kohdeorganisaation kaltaisen organisaation laajoihin ennustetarpeisiin nopeutensa ja skaalautuvuutensa ansiosta. Sen onnistunut ja laaja-alainen käyttöönotto edellyttää kuitenkin paitsi optimaalisen hienosäädön tason määrittämistä ja mahdollisten käyttökustannusten huomioimista, myös merkittäviä panostuksia käyttäjien koulutukseen, mallin toiminnan selittämiseen ja luottamuksen rakentamiseen. Power BI -alusta tarjoaa tähän hyvän pohjan, mutta tekoälyn "musta laatikko" -ongelma on ratkaistava käyttäjälähtöisesti.

## **5.2 Tutkimuksen merkitys**

Tämä tutkimus tarjoaa sekä käytännöllistä että tieteellistä lisäarvoa arvioimalla TimeGPT-mallin soveltuvuutta ja tehokkuutta taloudellisessa kustannusennustamisessa liikenteenohjauksen kaltaisessa monimutkaisessa toimintaympäristössä.

### **5.2.1 Käytännön merkitys kohdeorganisaatiolle**

Tutkimuksen tulokset tarjoavat kohdeorganisaatiolle konkreettisia suosituksia TimeGPT-mallin hyödyntämiseen ja ennusteprosessin kehittämiseen:

#### **Suosituksset TimeGPT:n käytöstä**

Tutkimus osoitti TimeGPT:n olevan erityisen tehokas suurten ja monimutkaisten aineistojen käsittelyssä, joissa perinteisten mallien suorituskyky heikkeni ja laskeutuminen-aika kasvoi merkittävästi. Suositeltavaa on hyödyntää TimeGPT:tä erityisesti näissä tilanteissa. Mallin hienosäätö organisaation omalla datalla paransi tarkkuutta, mutta optimaalisen kustannushyötysuhteen saavuttamiseksi suositellaan

maltillista hienosäätöä, sillä lisäaskeleet eivät tuottaneet merkittävää parannusta. Vaikka TimeGPT on laskennallisesti nopea, sen mahdollisia API-kustannuksia tulee arvioida suhteessa saavutettuihin hyötyihin.

### **Mahdolliset prosessimuutokset**

TimeGPT:n nopeus ja kyky käsitellä laajoja aineistoja mahdollistavat ennusteprosessin tehostamisen. Malli voi tuottaa luotettavan pohjaennusteen, mikä vapauttaa ennustajien aikaa manuaalisesta datan käsittelystä ja rutiiniennusteiden laatimisesta kohti analyyttisempää työtä, kuten poikkeamien syiden selvittämistä, skenaarioiden mallintamista ja ennusteiden laadullista rikastamista. Ennustekierroksia voidaan mahdollisesti tihentää tai tehdä iteratiivisemmin mallin nopeuden ansiosta.

### **Power BI:n rooli ja päätöksenteon tuki**

TimeGPT-ennusteiden integrointi Power BI -alustaan oli keskeistä työkalun käytettävyydelle. Ennustajien kysely vahvisti, että itse Power BI -raportti koettiin erittäin helppokäyttöiseksi ja sen ominaisuuksia, kuten visuaalista vertailua toteumaan ja budjettiin sekä datan porautumismahdollisuuksia, arvostettiin. Nämä ominaisuudet tukevat suoraan ennusteiden tulkintaa, poikkeamien analysointia ja siten päätöksentekoa. On kuitenkin tärkeää huomata kontrasti: vaikka Power BI -alusta sai kiitosta, se ei poistanut käyttäjien epävarmuutta itse TimeGPT-ennusteen logiikkaa kohtaan. Power BI toimii siis tehokkaana ja käyttäjäystävällisenä rajapintana tekoälyennusteiden esittämiseen ja hyödyntämiseen, mutta luottamuksen rakentaminen itse TimeGPT-malliin vaatii erillisiä toimenpiteitä liittyen sen läpinäkyvyyteen.

## 5.2.2 Tieteellinen merkitys

Tämä tutkimus edistää akateemista keskustelua monin tavoin:

### **Tutkimusaukon täyttäminen**

Suurten perustamallien (foundation models) kuten TimeGPT:n soveltamista nimenaan taloudelliseen kustannusennustamiseen organisaatiokontekstissa on tutkittu vähän. Tämä työ tarjoaa empiiristä näyttöä TimeGPT:n suorituskyvystä ja rajoituksista tällä spesifillä, käytännönläheisellä sovellusalueella.

### **Empiirinen vertailu ja kontekstin merkitys**

Tutkimus tarjoaa systemaattisen vertailun TimeGPT:n (sekä zero-shot että hienosäädettynä) ja vakiintuneiden aikasarjamallien välillä käyttäen laajaa, todellista ja osin epätäydellistä organisaation kustannusdataa. Tulokset korostavat, miten datan laajuus ja monimutkaisuus vaikuttavat mallien suhteelliseen suorituskykyyn, mikä on tärkeä havainto mallivalintaa koskevassa keskustelussa.

### **Hienosäädön käytännön vaikutukset**

Analysoimalla eri hienosäätöaskelmien vaikutusta tarkkuuteen ja laskenta-aikaan, tutkimus antaa käytännönläheistä tietoa perustamallien hienosäädön optimoinnista organisaatiokohtaisiin aikasarjoihin, osoittaen samalla pienenevän lisähyödyn ilmiö

### **Käyttäjäkokemuksen integrointi**

Käyttäjäkokemusten analyysi oli tärkeää tekoälyn käyttöönoton ymmärtämiseksi. Se paljasti keskeisen haasteen: vaikka Power BI -raportti koettiin helppokäyttöiseksi, itse TimeGPT-ennusteen "musta laatikko" -luonne eli toimintalogiikan epäselvyys nousi merkittäväksi esteeksi. Tämä heikensi luottamusta ja rajoitti laajempaa hyväksyntää, vaikka aktiivikäyttäjät näkivätkin mallissa jonkin verran hyötyä. Havainto korostaa, ettei tekninen tarkkuus yksin riitä, vaan laaja-alainen käyttöönotto edellyttää luottamusta, joka saavutetaan lisäämällä mallin läpinäkyvyyttä, selittämällä sen toimintaa ja tukemalla käyttäjiä.

Edellä mainittujen osioiden kautta tutkimus ei ainoastaan tarjoa käytännön ohjeita kohdeorganisaatiolle, vaan myös laajentaa tieteellistä ymmärrystä uusimpien tekoälymallien mahdollisuuksista, haasteista ja käyttöönnotosta talousennustamisen alalla.

### 5.3 Tutkimuksen rajoitukset

Tutkimuksen tulosten tulkinnassa ja soveltamisessa on tärkeää huomioida tietyt rajoitukset. Merkittävin rajoitus koskee tutkimuksen kohdetta ja aineistoa: analyysi keskittyy yksinomaan kohdeorganisaation kustannusdataan, jättäen ulkopuolelle muut taloudelliset muuttujat, kuten tulot ja kassavirrat, sekä ulkoiset markkinatekijät. Tämä tarkoittaa, että havainnot TimeGPT:n suorituskyvystä ovat vahvasti sidoksissa tähän nimenomaiseen organisaatioon ja datatyyppeihin, eikä tuloksia voida suoraan yleistää muihin yrityksiin, toimialoihin tai ennustekohteisiin ilman erillistä validointia.

Toinen huomionarvoinen seikka liittyy datan esikäsittelyyn, joka pidettiin tarkoituksellisen vähäisenä realistisen käyttöympäristön simuloimiseksi. Vaikka perusteellisempi datan puhdistus voisi mahdollisesti parantaa ennusteiden tarkkuutta, tämä tutkimus arvioi mallien toimintaa käytännönläheisemmällä, vähemmän käsitellyllä datalla. Tämä valinta vaikuttaa väistämättä mallien absoluuttiseen suorituskykyyn ja vertailtavuuteen. Lisäksi tutkimuksen tulokset ovat sidoksissa käytettyihin ohjelmistoversioihin ja erityisesti TimeGPT:n API-rajapintaan ja sen mahdollisiin kustannuksiin, mikä voi vaikuttaa tulosten toistettavuuteen ja mallin laajempaan käyttöönnottoon tulevaisuudessa.

Näistä rajoituksista huolimatta tutkimus tarjoaa kuitenkin perustellun kuvan TimeGPT-mallin potentiaalista ja haasteista kohdeorganisaation kustannusennustamisessa ja luo pohjaa jatkokehitykselle ja -tutkimukselle määritellyssä viitekehityksessään.

## 5.4 Jatkotutkimusaiheet

Laadittu tutkimus TimeGPT-mallin soveltuvuudesta kohdeorganisaation kustannusennustamiseen avaa useita jatkotutkimusaiheita. Alla esitellään keskeisimpiä jatkokehitysideoita, jotka perustuvat tutkimuksen havaintoihin ja tunnistettuihin rajoituksiin. Nämä jatkotutkimusaiheet auttaisivat paitsi syventämään ymmärrystä TimeGPT:n ja muiden tekoälymallien soveltuvuudesta taloudelliseen ennustamiseen, myös tukemaan kohdeorganisaatiota ennusteprosessien jatkuvassa kehittämisessä ja tehostamisessa.

### **Ulkoisten muuttujien (eksogeenisten tekijöiden) integrointi malliin**

Tässä tutkimuksessa keskityttiin mallien vertailuun pelkän historiallisen kustannusdatan perusteella, rajaten ulkoiset tekijät pois. Merkittävä jatkotutkimusaihe olisi selvittää, miten relevanttien eksogeenisten muuttujien, kuten yleisten talousindikaattoreiden, toimialakohtaisten trendien tai kohdeorganisaation oman operatiivisen datan (esim. tuotannontekijä määrien), sisällyttäminen vaikuttaisi TimeGPT:n ennustetarkkuuteen. Erityisen kiinnostavaa olisi tutkia, voidaanko kohdeorganisaation avointa dataa (esim. mittausasemien tietoja) hyödyntää ennusteiden tarkentamisessa, sekä voisiko TimeGPT:tä soveltaa myös itse liikennedatan ennustamiseen. Tämä vaatii kuitenkin huolellista validointia, sillä aikaisemmat tutkimukset ovat antaneet viitteitä siitä, etteivät eksogeeniset muuttujat aina paranna perustamallien suorituskykyä.

### **Vertailu uusiin tekoälymalleihin**

Aikasarjaennustamisen ala kehittyy nopeasti. Olisi hyödyllistä verrata TimeGPT:n suorituskykyä kohdeorganisaation datalla muihin uusiin ja kehittyneisiin tekoälymalleihin, erityisesti muihin transformer-pohjaisiin malleihin tai suuriin kielimalleihin (LLM), jotka on adaptoitu aikasarjatehtäviin. Tämä auttaisi sijoittamaan TimeGPT:n suorituskyvyn laajempaan "state-of-the-art" -kontekstiin.

### **Hybridimallien ja selitettävyyden kehittäminen**

Tulevaisuudessa voisi tutkia hybridimalleja, joissa yhdistetään TimeGPT:n vahvuuksia muihin menetelmiin. Yksi mielenkiintoinen suunta olisi käyttää esimerkiksi anomaliatunnistusmenetelmiä tunnistamaan TimeGPT:n tekemiä suuria ennustevirheitä ja hyödyntää suurta kielimallia (LLM) tuottamaan näistä virheistä tai poikkeamista luonnollisen kielen selityksiä ennustajille. Tämä voisi merkittävästi parantaa ennusteiden läpinäkyvyyttä, tulkittavuutta ja käyttäjien luottamusta järjestelmään.

### **Datan esikäsittelyn vaikutuksen tutkiminen**

Koska tässä työssä datan esikäsittely pidettiin vähäisenä, jatkotutkimuksessa voitaisiin systemaattisesti arvioida erilaisten ja perusteellisempien datanpuhdistus- ja muokkaustekniikoiden vaikutusta TimeGPT:n ja muiden mallien ennustetarkkuuteen kohdeorganisaation datalla.

## Lähteet

- Ali, I., Warraich, N. F., & Butt, K. (2024). Acceptance and use of artificial intelligence and AI-based applications in education: A meta-analysis and future direction. *Information Development*, advance online publication. <https://doi.org/10.1177/02666669241257206>
- Bontempi, G., Taieb, S. B., & Le Borgne, Y. A. (2013). Machine learning strategies for time series forecasting. Teoksessa M.-A. Aufaure & E. Zimányi (toim.), *Business intelligence: Second European Summer School, eBISS 2012, tutorial lectures* (Lecture Notes in Business Information Processing 138, s. 62–77). Springer [https://doi.org/10.1007/978-3-642-36318-4\\_3](https://doi.org/10.1007/978-3-642-36318-4_3)
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). Chapman & Hall/CRC.
- Christou, P. (2023). How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research? *The Qualitative Report*, 28(7), 1968–1980. <https://doi.org/10.46743/2160-3715/2023.6406>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Deforce, B., Baesens, B., & Serral Asensio, E. (2024). Leveraging time-series foundation models in smart agriculture for soil-moisture forecasting. *arXiv*. <https://arxiv.org/abs/2405.18913>

- Demir, H., & Karaođlan, K. (2024). A comparative performance analysis of LSTM autoencoder and TimeGPT models in time-series anomaly detection. ResearchGate. Noudettu 10.2.2025 osoitteesta [https://www.researchgate.net/publication/385301485 A Comparative Performance Analysis of LSTM Autoencoder and TimeGPT Models in Time Series Anomaly Detection](https://www.researchgate.net/publication/385301485_A_Comparative_Performance_Analysis_of_LSTM_Autoencoder_and_TimeGPT_Models_in_Time_Series_Anomaly_Detection)
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253-263. <https://doi.org/10.1080/07350015.1995.10524599>
- Eldele, E., Ragab, M., Chen, Z., Wu, M., & Li, X. (2024). TSLANet: Rethinking transformers for time-series representation learning. *arXiv*. <https://arxiv.org/abs/2404.08472>
- Fan, Y., Song, T., Feng, C., Song, K., Liu, C., & Jiang, D. (2024). Fine-tuning pre-trained large time series models for prediction of wind turbine SCADA data. *arXiv*. <https://arxiv.org/abs/2412.00403>
- Garza, A., Challu, C., & Mergenthaler-Canseco, M. (2024). TimeGPT-1. *arXiv*. <https://arxiv.org/abs/2310.03589>
- Gruver, N., Finzi, M., Qiu, S., & Wilson, A. G. (2023). Large language models are zero-shot time series forecasters. *arXiv*. <https://arxiv.org/abs/2310.07820>
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean-squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)

- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (3rd ed.). OTexts. Noudettu 22.1.2025 osoitteesta <https://otexts.com/fpp3>
- Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., ... Wen, Q. (2023). Time-LLM: Time series forecasting by reprogramming large language models. *arXiv*. <https://arxiv.org/abs/2310.01728>
- Kamalov, F., Zicmane, I., Safaraliev, M., Smail, L., Senyuk, M., & Matrenin, P. (2024). Attention-based load forecasting with bidirectional finetuning. *Energies*, 17(18), 4699. <https://doi.org/10.3390/en17184699>
- Liao, W., Wang, S., Yang, D., Yang, Z., Fang, J., Rehtanz, C., & Porté-Agel, F. (2025). TimeGPT in load forecasting: A large time series model perspective. *Applied Energy*, 379, 124973. <https://doi.org/10.1016/j.apenergy.2024.124973>
- Liu, H., Zhao, Z., Wang, J., Kamarthi, H., & Prakash, B. A. (2024). LSTPrompt: Large language models as zero-shot time series forecasters by long-short-term prompting. *arXiv*. <https://arxiv.org/abs/2402.16132>
- Merrill, M. A., Tan, M., Gupta, V., Hartvigsen, T., & Althoff, T. (2024). Language models still struggle to zero-shot reason about time series. *arXiv*. <https://arxiv.org/abs/2404.11757>
- Nixtlaverse. (2024). *TimeGPT*. Nixtla. Noudettu 10.1.2025 osoitteesta <https://nixtlaverse.nixtla.io>

- Paroha, A. D., & Chotrani, A. (2024). A comparative analysis of TimeGPT and Time-LLM in predicting ESP maintenance needs in the oil and gas sector. *International Journal of Computer Applications*, 186(8), 1–10. [https://www.researchgate.net/publication/378550404\\_A\\_Comparative\\_Analysis\\_of\\_TimeGPT\\_and\\_Time-LLM\\_in\\_Predicting\\_ESP\\_Maintenance\\_Needs\\_in\\_the\\_Oil\\_and\\_Gas\\_Sector](https://www.researchgate.net/publication/378550404_A_Comparative_Analysis_of_TimeGPT_and_Time-LLM_in_Predicting_ESP_Maintenance_Needs_in_the_Oil_and_Gas_Sector)
- Romero-Brufau, S., Whitford, D., Johnson, M. G., Hickman, J., Morlan, B. W., Therneau, T., Naessens, J., & Huddleston, J. M. (2021). Using machine learning to improve the accuracy of patient deterioration predictions: Mayo Clinic Early Warning Score (MC-EWS). *Journal of the American Medical Informatics Association*, 28(6), 1207–1215. <https://doi.org/10.1093/jamia/ocaa347>
- Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review (2005–2019). *Applied Soft Computing*, 90, 106181. <https://doi.org/10.1016/j.asoc.2020.106181>
- Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.
- Song, Z., Lu, Q., Xu, H., Zhu, H., Buckeridge, D. L., & Li, Y. (2024). TimelyGPT: Extrapolatable transformer pre-training for long-term time-series forecasting in healthcare. *arXiv*. <https://arxiv.org/abs/2312.00817>
- Spiliotis, E. (2023). Time series forecasting with statistical, machine learning, and deep learning methods: Past, present, and future. Teoksessa S. Makridakis (toim.), *Forecasting with artificial intelligence* (s. 49–75). Springer. [http://dx.doi.org/10.1007/978-3-031-35879-1\\_3](http://dx.doi.org/10.1007/978-3-031-35879-1_3)

- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. Teoksessa M. J. Metzger & A. J. Flanagin (toim.), *Digital media, youth, and credibility* (s. 73–100). MIT Press. <https://search.is-suelab.org/resource/the-main-model-a-heuristic-approach-to-understanding-technology-effects-on-credibility.html>
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45 <https://doi.org/10.1080/00031305.2017.1380080>
- Taylor, S. J., & Letham, B. (n.d.). *Prophet* (versio 1.1) [Tietokoneohjelmisto]. Noudettu 10.1.2025 osoitteesta <https://facebook.github.io/prophet/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. <https://arxiv.org/abs/1706.03762>
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term time-series forecasting. *Advances in Neural Information Processing Systems*, 34, 22419–22430. <https://arxiv.org/abs/2106.13008>

Zhang, H., Arvin, C., Efimov, D., Mahoney, M. W., Perrault-Joncas, D., Ramasubramanian, S., Wilson, A. G., & Wolff, M. (2024). LLMForecaster: Improving seasonal event forecasts with unstructured textual data. *arXiv*. <https://arxiv.org/abs/2412.02525>

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Teoksessa Proceedings of the 35th AAAI Conference on Artificial Intelligence* (Vol. 35). AAAI Press. <https://arxiv.org/abs/2012.07436>

## Liitteet

### Liite 1. DM-HLN testi

```

import pandas as pd
import numpy as np
from scipy.stats import norm
import itertools

def dm_test_hln(errors_model_1, errors_model_2, h=1):
    """
    Diebold-Mariano-testi Harvey, Leybourne ja Newbold (HLN) -korjauksella.

    Parametrit:
    -----
    errors_model_1 : array-like
        Mallin 1 ennustevirheet (toteuma - ennuste).
    errors_model_2 : array-like
        Mallin 2 ennustevirheet (toteuma - ennuste).
    h : int
        Ennustehorisontti (askelmäärä). Tässä oletetaan h=1.

    Palauttaa:
    -----
    DM_HLN_stat : float
        HLN-korjattu DM-testisuure.
    p_value : float
        Testin p-arvo.
    """
    # Muodostetaan erotuslossi: käytetään MSE-erotusta tässä esimerkissä
    d = (errors_model_1**2 - errors_model_2**2)
    T = len(d)

    # Lasketaan erotuslossien keskiarvo
    d_bar = np.mean(d)
    # Lasketaan erotuslossien varianssi
    var_d = np.var(d, ddof=1)
    # Lasketaan DM-testisuure
    DM_stat = d_bar / np.sqrt(var_d/T)

    # HLN-korjaus DM-testisuureelle
    factor = np.sqrt((T+1 - 2*h + (h*(h-1))/T) / T)
    DM_HLN_stat = DM_stat * factor

    # Lasketaan p-arvo normaalijakauman perusteella
    p_value = 2*(1 - norm.cdf(np.abs(DM_HLN_stat)))

    return DM_HLN_stat, p_value

if __name__ == "__main__":
    # Lue ennustedata CSV-tiedostosta
    df = pd.read_csv('forecasts.csv', parse_dates=['date'])

```

```
# Ennustemallien sarakenimet
model_cols = ['zero_shot', '10_askelta', '30_askelta', '50_askelta', 'ETS', 'SARIMA', 'Prophet']

# Lasketaan jokaisen mallin ennustevirheet
for m in model_cols:
    df[f'error_{m}'] = df['actual'] - df[m]

# Luodaan lista virhesarakkeista
error_cols = [f'error_{m}' for m in model_cols]

# Käydään läpi kaikki mahdolliset malliparit
# itertools.combinations tuottaa kaikki mahdolliset 2-alkioiset yhdistelmät ilman toistoa
for (m1, m2) in itertools.combinations(error_cols, 2):
    e1 = df[m1].values
    e2 = df[m2].values
    DM_HLN_stat, p_value = dm_test_hln(e1, e2, h=1)

# Tulostetaan testin tulokset
# Mallin nimet haetaan virhesarakkeen nimen kautta, esim. error_zero_shot -> zero_shot
model_name_1 = m1.replace('error_', '')
model_name_2 = m2.replace('error_', '')
print(f"DM-HLN-testi ({model_name_1} vs {model_name_2}):")
print("Testisuure:", DM_HLN_stat)
print("p-arvo:", p_value)
print("-"*50)
```

## Liite 2. Ajallisen cross-validationin toteutus MAE-mittarilla

```

from nixtla import NixtlaClient
import pandas as pd
import numpy as np

# Alusta NixtlaClient käyttämällä omaa API-avainta
nixtla_client = NixtlaClient(
    api_key='nixtla-tok-xxxxx'
)

nixtla_client.validate_api_key()

# Tiedostopolku
file_path = 'Ennustedata Kaikki.csv'

# Lue CSV-tiedosto
df = pd.read_csv(file_path, sep=';', decimal=',', encoding='latin1')

# Siistitään sarakenimet
df.columns = [col.strip() for col in df.columns]

# Tunnistetaan aikaleimasarake (oletetaan, että se on ensimmäinen sarake)
timestamp_column = df.columns[0]
df[timestamp_column] = pd.to_datetime(df[timestamp_column], format='%Y%m')

# Korvataan puuttuvat tai nolla-arvot datassa arvolla 0.0
df = df.fillna(0.0) # Korvataan NaN-arvot arvolla 0.0
df = df.replace(0, 0.0) # Varmistetaan, että nolla-arvot on eksplisiittisesti 0.0

# Muotoillaan DataFrame pitkään muotoon
df_long = df.melt(id_vars=[timestamp_column], var_name='unique_id', value_name='value')
df_long['value'] = df_long['value'].fillna(0.0) # Varmistetaan, että NaN-arvoja ei jää pitkään muotoon

# Parametrit
h = 10 # Ennustehorisontti
freq = 'MS' # Kuukausitasoinen data
finetune_steps = 50 # Hienosäätöaskeleet
unique_ids = df_long['unique_id'].unique() # Sarjojen tunnisteet

dates = df_long[timestamp_column].unique() # Aikaleimat
dates = sorted(dates)

cutoff_count = 3 # Cross-validation-jaksojen määrä
valid_cutoffs = [d for d in dates if d < dates[-h]] # Rajaus: vain ennen viimeistä ennustejaksoa

start_date = dates[0]
# Lasketaan kuukausien lukumäärä suhteessa aloituspäivämäärään
valid_cutoffs = [d for d in valid_cutoffs if ((d.year - start_date.year)*12 + (d.month - start_date.month)) >=
36]

# Päivitetään cutoff-jaksojen määrä, jos validit rajaukset jäävät alle määritetyn rajan
if len(valid_cutoffs) < cutoff_count:

```

```

cutoff_count = len(valid_cutoffs)

if cutoff_count == 0:
    print("Ei löytynyt sopivia cutoff-päiviä, joilla olisi vähintään 36 kuukautta dataa ja ennustehorisontti saavilla.")
    exit()

# Valitaan tasaisesti sijoittuvat cutoff-päivät
cutoff_indices = np.linspace(0, len(valid_cutoffs)-1, cutoff_count, dtype=int)
cutoff_dates = [valid_cutoffs[i] for i in cutoff_indices]

mae_list = [] # Lista MAE-arvojen tallentamiseen

for cutoff_date in cutoff_dates:
    print(f"Käsitellään cutoff-päivää: {cutoff_date}")
    train_df = df_long[df_long[timestamp_column] <= cutoff_date]

    # Tarkistetaan, että jokaisessa sarjassa on vähintään 36 havaintoa
    counts = train_df.groupby('unique_id').size()
    valid_ids = counts[counts >= 36].index
    train_df = train_df[train_df['unique_id'].isin(valid_ids)]

    if len(valid_ids) == 0:
        print(f"Ei sarjoja, joissa >=36 havaintoa cutoff-päivänä {cutoff_date}, ohitetaan tämä jakso.")
        continue

    # Ennusteiden luominen
    fcst_df = nixtla_client.forecast(
        df=train_df,
        h=h,
        freq=freq,
        time_col=timestamp_column,
        target_col='value',
        id_col='unique_id',
        finetune_steps=finetune_steps
    )

    forecast_horizon_end = cutoff_date + pd.DateOffset(months=h)

    # Suodatetaan toteutunut data ennusteen vertailua varten
    actual_df = df_long[(df_long[timestamp_column] > cutoff_date) & (df_long[timestamp_column] <=
forecast_horizon_end)]

    # Yhdistetään ennusteet ja toteutumat
    fcst_df = fcst_df.rename(columns={'TimeGPT': 'forecast'})
    merged = pd.merge(fcst_df, actual_df, on=[timestamp_column, 'unique_id'], how='inner')

    if merged.empty:
        print(f"Ei päällekkäistä toteutunutta dataa cutoff-päivälle {cutoff_date}, ohitetaan.")
        continue

    # Lasketaan absoluuttinen virhe ja MAE
    merged['abs_error'] = (merged['forecast'] - merged['value']).abs()
    mae = merged['abs_error'].mean()
    mae_list.append(mae)

```

```
print(f"MAE cutoff-päivälle {cutoff_date}: {mae}")

if mae_list:
    # Lasketaan keskimääräinen MAE cross-validation-jaksojen yli
    cv_mae = np.mean(mae_list)
    print(f"Cross-validated MAE (keskiarvo {len(mae_list)} jaksosta): {cv_mae}")
else:
    print("Ei käsitelty yhtään validia jaksoa datarajoitteiden vuoksi.")

print("Cross-validation valmis.")
```

### Liite 3. SARIMA ennuste

```

import pandas as pd
import warnings
import pmdarima as pm
from sklearn.metrics import mean_absolute_error, mean_squared_error
import sys
import os

# Piilota varoitukset siistin tulostuksen vuoksi
warnings.filterwarnings("ignore")

# Varmista työskentelyhakemisto
current_dir = os.getcwd()
print(f"Nykyinen työskentelyhakemisto: {current_dir}")

# Määritä oikea polku CSV-tiedostolle
file_path = 'Ennustedata Raide.csv'

# Tarkista, että tiedosto on olemassa
if not os.path.isfile(file_path):
    print(f"Tiedostoa ei löytynyt: {file_path}")
    sys.exit(1)

# Lue CSV-tiedosto
try:
    df = pd.read_csv(file_path, sep=';', decimal=',', encoding='latin1')
    print("CSV-tiedosto ladattu onnistuneesti.")
except Exception as e:
    print(f"Virhe CSV-tiedoston lukemisessa: {e}")
    sys.exit(1)

# Siisti sarakenimet
df.columns = [col.strip() for col in df.columns]
print("Sarakenimet siistitty.")

# Tunnista aikaleimasarake (oletetaan, että se on ensimmäinen sarake)
timestamp_column = df.columns[0]
try:
    df[timestamp_column] = pd.to_datetime(df[timestamp_column], format='%Y%m')
    print(f"Aikaleimasarake '{timestamp_column}' muunnettu datetime-muotoon.")
except Exception as e:
    print(f"Virhe aikaleimasarakkeen muuntamisessa: {e}")
    sys.exit(1)

# Korvaa puuttuvat tai nolla-arvot arvolla 0.0
df = df.fillna(0.0) # Korvaa NaN-arvot arvolla 0.0

# Korvaa mahdolliset merkkijonot '0' arvolla 0.0 ja poista tuhaterotitimet
for col in df.columns[1:]:
    df[col] = df[col].astype(str).str.replace(' ', '').str.replace(',', '').replace('0', '0.0')

print("Puuttuvat ja nolla-arvot korvattu arvolla 0.0.")

```

```

# Muunna sarakkeet numeerisiksi, coercing virheet NaN-arvoiksi ja täytä ne
for col in df.columns[1:]:
    df[col] = pd.to_numeric(df[col], errors='coerce').fillna(0.0)

print("Numeeriset sarakkeet muunnettu liukuluvuiksi.")

# Tarkista päällekkäiset sarakkeet
original_columns = df.columns.tolist()
df = df.loc[:, ~df.columns.duplicated()]
if len(original_columns) != len(df.columns):
    print("Päällekkäiset sarakkeet löytyivät ja poistettiin.")

# Aseta aikaleima indeksiksi
df.set_index(timestamp_column, inplace=True)
print("Aikaleima asetettu indeksiksi.")

# Luo täydellinen aikaväli ja lisää puuttuvat kuukaudet arvoilla 0.0
full_time_index = pd.date_range(start=df.index.min(), end=df.index.max(), freq='MS')
missing_dates = full_time_index.difference(df.index)
if not missing_dates.empty:
    print(f"Puuttuvat päivämäärät datassa: {missing_dates}")
    missing_df = pd.DataFrame(index=missing_dates, columns=df.columns).fillna(0.0)
    df = pd.concat([df, missing_df]).sort_index()
    print("Puuttuvat päivämäärät lisätty arvoilla 0.0.")

# Ennustehorisontti
h = 10 # Ennustehorisontti (10 kuukautta)

# Alusta tyhjä sanakirja ennusteiden tallentamiseen
forecast_dict = {}

for unique_id in df.columns:
    series = df[unique_id]

    # Tarkista, että sarjassa on riittävästi havaintoja
    if len(series) < 24: # Esim. vähintään kaksi kausittaista sykliä
        print(f" Ei tarpeeksi dataa ennustamiseen {unique_id}:lle. Ohitetaan.")
        continue

    try:
        # Sovita auto_arima-malli koko aikasarjaan
        model = pm.auto_arima(
            series,
            seasonal=True,
            m=12, # Kausivaihtelun periodi (kuukausittainen data)
            trace=False,
            error_action='ignore',
            suppress_warnings=True,
            stepwise=True
        )

        # Ennusta
        forecast = model.predict(n_periods=h)

```

```

forecast_index = pd.date_range(start=series.index[-1] + pd.DateOffset(months=1), periods=h,
freq='MS')
forecast_series = pd.Series(forecast, index=forecast_index)

# Tallenna ennuste sanakirjaan
forecast_dict[unique_id] = forecast_series

print(f"Ennuste valmis {unique_id}:lle auto_arimalla.")
except Exception as e:
print(f"Virhe ennustettaessa {unique_id} auto_arimalla: {e}")
# Yritä manuaalisesti määritellyillä ARIMA-parametreilla
try:
# Luo ja sovita manuaalinen ARIMA-malli
manual_model = pm.ARIMA(order=(1,1,1), seasonal_order=(1,1,0,12))
manual_model.fit(series)

# Ennusta
forecast = manual_model.predict(n_periods=h)
forecast_index = pd.date_range(start=series.index[-1] + pd.DateOffset(months=1), periods=h,
freq='MS')
forecast_series = pd.Series(forecast, index=forecast_index)

# Tallenna ennuste
forecast_dict[unique_id] = forecast_series

print(f"Ennuste valmis {unique_id}:lle manuaalisella ARIMA:lla.")
except Exception as e2:
print(f"Virhe ennustettaessa {unique_id} manuaalisella ARIMA:lla: {e2}")
# Voit lisätä muita vaihtoehtoja tarvittaessa

# Varmista, että forecast_dict sisältää ennusteita
if not forecast_dict:
print("Ennusteita ei tuotettu.")
sys.exit(1)

# Muunna ennustesanakirja DataFrameksi
forecast_df = pd.DataFrame(forecast_dict)
forecast_df.reset_index(inplace=True)
forecast_df.rename(columns={'index': 'timestamp_column'}, inplace=True)
print("Ennustesanakirja muunnettu DataFrameksi.")

# Yhdistä historiallinen data ja ennusteet
historical_df = df.reset_index()

# Lisää ennustedata
combined_df = pd.concat([historical_df, forecast_df], ignore_index=True)
print("Historiallinen data ja ennusteet yhdistetty.")

# Järjestä aikaleiman mukaan
final_combined_df = combined_df.sort_values(by='timestamp_column')
print("Yhdistetty data järjestetty aikaleiman mukaan.")

# Varmista, että sarakkeiden järjestys vastaa alkuperäistä DataFramea
try:
final_combined_df = final_combined_df[[timestamp_column] + list(df.columns)]

```

```
    print("Sarakkeiden järjestys vastaa alkuperäistä DataFramea.")
except KeyError as e:
    print(f"Virhe sarakkeiden järjestämisessä: {e}")
    sys.exit(1)

# Tallenna CSV-tiedostoon
output_file = 'historical_and_forecast_combined_pivot_SARIMA_auto.csv'
try:
    final_combined_df.to_csv(output_file, index=False)
    print(f"Data käsitelty ja SARIMA (auto_arima) -ennuste tallennettu tiedostoon '{output_file}'.")
except Exception as e:
    print(f"Virhe CSV-tiedoston tallentamisessa: {e}")
    sys.exit(1)
```

## Liite 4. Prophet ennuste

```

import pandas as pd
import warnings
from prophet import Prophet
import sys

# Piilota varoitukset siistin tulostuksen vuoksi
warnings.filterwarnings("ignore")

# Tiedostopolut
input_file_path = 'Ennustedata Kaikki.csv' # Syötetiedoston polku
output_forecast = 'Forecasts Kaikki.csv' # Ennustetiedoston polku

# Lue CSV-tiedosto
df = pd.read_csv(input_file_path, sep=';', decimal=',', encoding='latin1')

# Siisti sarakenimet
df.columns = [col.strip() for col in df.columns]

# Tunnista aikaleimasarake (oletetaan, että se on ensimmäinen sarake)
timestamp_column = df.columns[0]
df[timestamp_column] = pd.to_datetime(df[timestamp_column], format='%Y%m')

# Korvaa puuttuvat tai nolla-arvot arvolla 0.0
df = df.fillna(0.0)
df = df.replace(0, 0.0)

# Aseta aikaleima indeksiksi
df.set_index(timestamp_column, inplace=True)

# Ennustehorisontti
h = 10 # Ennustehorisontti (10 kuukautta)

# Alusta DataFrame kaikkien ennusteiden keräämiseen
all_forecasts = pd.DataFrame()

# Iteroi kaikkien aikasarjojen yli
for unique_id in df.columns:
    series = df[unique_id]

    # Tarkista, että sarjassa on riittävästi havaintoja (esim. vähintään 24 kuukautta)
    if len(series) < 24:
        print(f"Ei tarpeeksi dataa ennustamiseen {unique_id}:lle. Ohitetaan.")
        continue

    try:
        # Valmistele data Prophet-mallia varten
        prophet_df = pd.DataFrame({
            'ds': series.index, # Aikaleimat
            'y': series.values # Arvot
        })

```

```

# Alusta Prophet-malli
model = Prophet(
    yearly_seasonality=True, # Kausivaihtelu vuositasolla
    weekly_seasonality=False, # Ei viikkotason kausivaihtelua
    daily_seasonality=False # Ei päivittäistä kausivaihtelua
)

# Sovita Prophet-malli
model.fit(prophet_df)

# Luo tulevaisuuden DataFrame
future = model.make_future_dataframe(periods=h, freq='MS') # Kuukausittainen data

# Tee ennuste
forecast = model.predict(future)

# Ota ennustetut arvot talteen
forecast_series = forecast.set_index('ds')['yhat'].tail(h)

print(f"Ennuste valmis {unique_id}:lle.")
print(forecast_series)

# Lisää ennuste all_forecasts-DataFrameeseen
all_forecasts[unique_id] = forecast_series.values
if all_forecasts.empty:
    all_forecasts.index = forecast_series.index
else:
    # Varmista, että indeksit täsmäävät
    if not all_forecasts.index.equals(forecast_series.index):
        all_forecasts.index = forecast_series.index

except Exception as e:
    print(f"Virhe ennustettaessa {unique_id}: {e}")

# Valmistele yhdistetty DataFrame, kun kaikki ennusteet on tehty
if not all_forecasts.empty:
    # Aseta ennustepäivämäärät indeksiksi
    all_forecasts.index = forecast_series.index

    # Nollaa indeksi ja muuta päivämäärät sarakkeeksi
    all_forecasts.reset_index(inplace=True)

    # Muotoile päivämäärä YYYY-MM-muotoon ennen uudelleennimeämistä
    all_forecasts['ds'] = all_forecasts['ds'].dt.strftime('%Y-%m')

    # Nimeä 'ds'-sarake selkeämmäksi (esim. 'Päivämäärä')
    all_forecasts.rename(columns={'ds': 'Päivämäärä'}, inplace=True)

    # Tallenna ennusteet CSV-tiedostoon
    all_forecasts.to_csv(output_forecast, index=False, sep=';', decimal=',')

    print(f"Kaikki ennusteet on tallennettu tiedostoon {output_forecast}.")
else:
    print("Yhtään ennustetta ei tuotettu.")

```

## Liite 5. ETS ennuste

```

import pandas as pd
from statsmodels.tsa.holtwinters import ExponentialSmoothing
import warnings

# Piilota varoitukset siistin tulostuksen vuoksi
warnings.filterwarnings("ignore")

# Tiedoston sijainti
file_path = 'Ennustedata.csv'

# Lue CSV-tiedosto
df = pd.read_csv(file_path, sep=';', decimal=',', encoding='latin1')

# Siisti sarakenimet
df.columns = [col.strip() for col in df.columns]

# Tunnista aikaleimasarake (oletetaan, että se on ensimmäinen sarake)
timestamp_column = df.columns[0]
df[timestamp_column] = pd.to_datetime(df[timestamp_column], format='%Y%m')

# Korvaa puuttuvat tai nolla-arvot arvolla 0.0
df = df.fillna(0.0) # Korvaa puuttuvat arvot (NaN) arvolla 0.0
df = df.replace(0, 0.0) # Varmista, että nolla-arvot ovat eksplisiittisesti 0.0

# Aseta aikaleima indeksiksi
df.set_index(timestamp_column, inplace=True)

# Ennustehorisontti
h = 10 # Ennustehorisontti (10 kuukautta)

# Alusta tyhjä sanakirja ennusteiden tallentamiseen
forecast_dict = {}

# Iteroi jokaisen sarakkeen yli (jokainen sarake vastaa eri aikasarjaa)
for unique_id in df.columns:
    series = df[unique_id]

    # Tarkista, että sarjassa on riittävästi havaintoja
    if len(series) < 2:
        print(f" Ei tarpeeksi dataa ennustamiseen {unique_id}:lle. Ohitetaan.")
        continue

    try:
        # Sovita ETS-malli (Exponential Smoothing)
        model = ExponentialSmoothing(
            series,
            trend='add', # Additiivinen trendi
            seasonal='add', # Additiivinen kausivaihtelu
            seasonal_periods=12 # Oletetaan kuukausittainen data, jossa vuosittainen kausivaihtelu
        ).fit()

```

```
# Tee ennuste
forecast = model.forecast(h)

# Tallenna ennuste sanakirjaan
forecast_dict[unique_id] = forecast
except Exception as e:
    print(f"Virhe ennustettaessa {unique_id}: {e}")

# Muunna ennustesanakirja DataFrameksi
forecast_df = pd.DataFrame(forecast_dict)
forecast_df.reset_index(inplace=True)
forecast_df.rename(columns={'index': timestamp_column}, inplace=True)

# Yhdistä historiallinen ja ennustettu data
# Palauta indeksi, jotta yhdistäminen on mahdollista
historical_df = df.reset_index()

# Lisää ennustedata
combined_df = pd.concat([historical_df, forecast_df], ignore_index=True)

# Järjestä aikaleiman mukaan varmistaaksesi oikea järjestys
final_combined_df = combined_df.sort_values(by=timestamp_column)

# Valinnainen: Varmista, että sarakkeiden järjestys vastaa alkuperäistä DataFramea
original_columns = [timestamp_column] + [col for col in df.columns]
final_combined_df = final_combined_df[original_columns]

# Tallenna CSV-tiedostoon
final_combined_df.to_csv('historical_and_forecast_combined_ETS.csv', index=False)

print("Data käsitelty ja ETS-ennuste tallennettu onnistuneesti.")
```

## Liite 6. TimeGPT zero-shot ennuste

```

from nixtla import NixtlaClient
import pandas as pd

# Alusta NixtlaClient käyttämällä omaa API-avainta
nixtla_client = NixtlaClient(
    api_key='nixtla-tok-xxxxx'
)

# Varmista API-avaimen toimivuus
nixtla_client.validate_api_key()

# Tiedoston sijainti
file_path = 'Ennustedata Kaikki.csv'

# Lue CSV-tiedosto
df = pd.read_csv(file_path, sep=';', decimal=',', encoding='latin1')

# Siisti sarakenimet
df.columns = [col.strip() for col in df.columns]

# Tunnista aikaleimasarake (oletetaan, että se on ensimmäinen sarake)
timestamp_column = df.columns[0]
df[timestamp_column] = pd.to_datetime(df[timestamp_column], format='%Y%m')

# Korvaa puuttuvat tai nolla-arvot arvolla 0.0
df = df.fillna(0.0) # Korvaa puuttuvat arvot (NaN) arvolla 0.0
df = df.replace(0, 0.0) # Varmista, että nolla-arvot ovat eksplisiittisesti 0.0

# Muotoile DataFrame pitkään muotoon
df_long = df.melt(id_vars=[timestamp_column], var_name='unique_id', value_name='value')
df_long['value'] = df_long['value'].fillna(0.0) # Varmista, ettei pitkään muotoon jää NaN-arvoja

# Tee ennuste TimeGPT:n avulla ilman hienosäätöä
timegpt_fcst_df = nixtla_client.forecast(
    df=df_long,
    h=10, # Ennustehorisontti (10 kuukautta)
    freq='MS', # Kuukausitaso, kuukausien alku
    time_col=timestamp_column,
    target_col='value',
    id_col='unique_id'
)

# Yhdistä historiallinen ja ennustettu data
historical_df = df_long.rename(columns={'value': 'value'})
forecast_df = timegpt_fcst_df.rename(columns={'TimeGPT': 'value'})
# Poista historiadata ennustetusta datasta
forecast_df = forecast_df[~forecast_df[timestamp_column].isin(historical_df[timestamp_column])]
# Yhdistä historiallinen ja ennustettu data yhdeksi DataFrameeksi
combined_df = pd.concat([historical_df, forecast_df], axis=0, ignore_index=True)

# Pivotoi yhdistetty data siten, että unique_id-arvot ovat sarakkeina, säilytä alkuperäinen sarakejärjestys

```

```
final_combined_df = combined_df.pivot(index=timestamp_column, columns='unique_id', values='value').reset_index()
final_combined_df = final_combined_df[df.columns] # Varmista, että sarakkeiden järjestys vastaa alkuperäistä
final_combined_df.to_csv('historical_and_forecast_combined_pivot.csv', index=False)

print("Data käsitelty ja ennuste tallennettu onnistuneesti.")
```

## Liite 7. TimeGPT hienosäädetty ennuste

```

from nixtla import NixtlaClient
import pandas as pd

# Alusta NixtlaClient käyttämällä omaa API-avainta
nixtla_client = NixtlaClient(
    api_key='nixtla-tok-xxxxx'
)

# Varmista API-avaimen toimivuus
nixtla_client.validate_api_key()

# Tiedoston sijainti
file_path = 'Ennustedata Kaikki.csv'

# Lue CSV-tiedosto
df = pd.read_csv(file_path, sep=';', decimal=',', encoding='latin1')

# Siisti sarakenimet
df.columns = [col.strip() for col in df.columns]

# Tunnista aikaleimasarake (oletetaan, että se on ensimmäinen sarake)
timestamp_column = df.columns[0]
df[timestamp_column] = pd.to_datetime(df[timestamp_column], format='%Y%m')

# Korvaa puuttuvat tai nolla-arvot arvolla 0.0
df = df.fillna(0.0) # Korvaa NaN-arvot arvolla 0.0
df = df.replace(0, 0.0) # Varmista, että nolla-arvot ovat eksplisiittisesti 0.0

# Muotoile DataFrame pitkään muotoon
df_long = df.melt(id_vars=[timestamp_column], var_name='unique_id', value_name='value')
df_long['value'] = df_long['value'].fillna(0.0) # Varmista, ettei pitkään muotoon jää NaN-arvoja

# Tee ennuste TimeGPT:n avulla käyttäen hienosäätöä
timegpt_fcst_df = nixtla_client.forecast(
    df=df_long,
    h=10, # Ennustehorisontti (10 kuukautta)
    freq='MS', # Kuukausitaso, kuukausien alku
    time_col=timestamp_column,
    target_col='value',
    id_col='unique_id',
    finetune_steps=40 # Hienosäätöaskeleet
)

# Yhdistä historiallinen ja ennustettu data
historical_df = df_long.rename(columns={'value': 'value'})
forecast_df = timegpt_fcst_df.rename(columns={'TimeGPT': 'value'})
# Poista historiadata ennustetusta datasta
forecast_df = forecast_df[~forecast_df[timestamp_column].isin(historical_df[timestamp_column])]
# Yhdistä historiallinen ja ennustettu data yhdeksi DataFrameeksi
combined_df = pd.concat([historical_df, forecast_df], axis=0, ignore_index=True)

```

```
# Pivotoi yhdistetty data siten, että unique_id-arvot ovat sarakkeina, säilytä alkuperäinen sarakejärjestys
final_combined_df = combined_df.pivot(index=timestamp_column, columns='unique_id', values='value').reset_index()
final_combined_df = final_combined_df[df.columns] # Varmista, että sarakkeiden järjestys vastaa alkuperäistä
final_combined_df.to_csv('historical_and_forecast_combined_pivot.csv', index=False)

print("Data käsitelty ja ennuste tallennettu onnistuneesti.")
```

## Liite 8. Kysely ennustajille

# Power BI ennusteraportin sekä TimeGPT-ennusteen käyttökokemus ja hyödyllisyys

Tämä kyselylomake on tarkoitettu keräämään tietoa Power BI -ennusteraportin sekä TimeGPT-ennustemallin käyttökokemuksesta ja sen hyödyllisyydestä ennusteprosessissa.

\* Pakollinen

1. Mikä on roolisi ennusteprosessissa? \*

- Yritystason ennustaja
- Vastuualuekohtainen ennustaja
- Projektikohtainen ennustaja
- Muu

2. Hyödynsitkö ennusteen tekemisessä ennusteraporttia ja\tai TimeGPT ennustetta? \*

- Kyllä - Vain ennusteraporttia
- Kyllä - Ennusteraporttia sekä TimeGPT-ennustetta
- Muu

3. Kuinka helppokäyttöiseksi koit ennusteraportin käytön?



4. Mitkä ominaisuudet koit erityisen hyödyllisiksi? (Valitse kaikki soveltuvat)

- Graafinen esitystapa
- Porautumismahdollisuus eri tasoille (yritys, vastuualue, projekti)
- Suodatusvaihtoehdot
- Tiedonsiirto muihin työkaluihin
- Muu, mikä? \_\_\_\_\_

5. Mitä parannuksia toivoisit ennusteraporttiin tai TimeGPT-ennusteeseen liittyen?

6. Kuinka tarkaksi ja luotettavaksi koit TimeGPT ennustemallin tuottamat ennusteet?



7. Kuinka hyödylliseksi koit ennusteraportin Navita-ennusteen päivittämisessä?



8. Kuinka hyödylliseksi koit TimeGPT-ennusteen Navita-ennusteen päivittämisessä?



9. Miten TimeGPT-ennusteet vaikuttivat ennusteen tai resurssien suunnitteluun?

10. Mitä lisäkoulutusta tai tukea koet tarvitsevasi ennusteraporttiin tai TimeGPT-ennustemalliin liittyen?

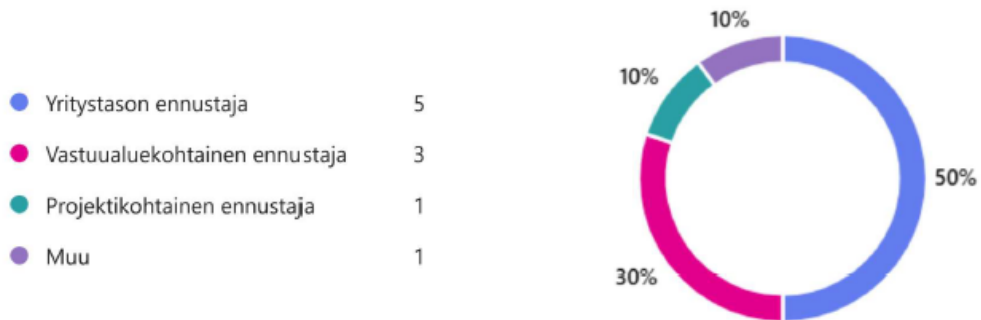
11. Onko sinulla muita huomioita tai kehitysehdotuksia ennusteraporttiin tai TimeGPT-ennustemalliin liittyen?

12. Suositteletko Power BI ennusteraportin ja TimeGPT-ennusteen hyödyntämistä työkaverillesi? (1=en, 3=ehkä, 5=kyllä)

1	2	3	4	5
---	---	---	---	---

## Liite 9. Kyselyn vastaukset

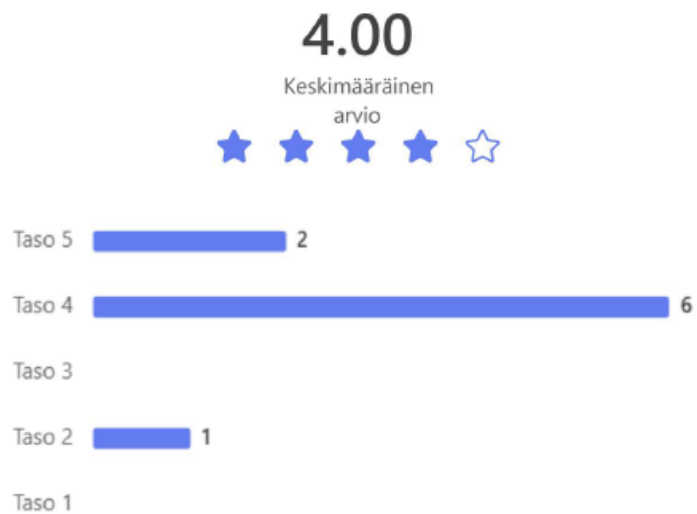
1. Mikä on roolisi ennusteprosessissa?



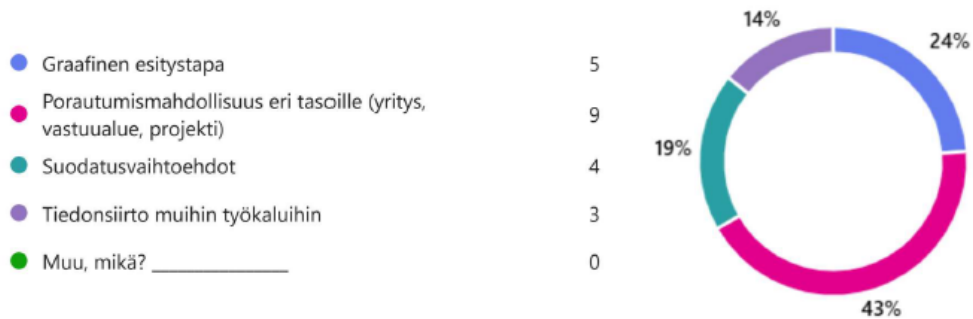
2. Hyödynsitkö ennusteen tekemisessä ennusteraporttia ja\tai TimeGPT ennustetta?



3. Kuinka helppokäyttöiseksi koit ennusteraportin käytön?



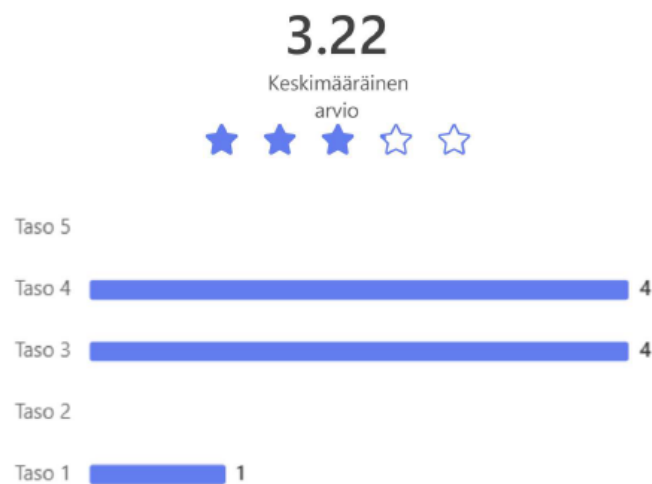
4. Mitkä ominaisuudet koit erityisen hyödyllisiksi? (Valitse kaikki soveltuvat)



5. Mitä parannuksia toivoisit ennusteraporttiin tai TimeGPT-ennusteeseen liittyen?

**4**  
Vastaukset

6. Kuinka tarkaksi ja luotettavaksi koit TimeGPT ennustemallin tuottamat ennusteet?



7. Kuinka hyödylliseksi koit ennusteraportin Navita-ennusteen päivittämisessä?



8. Kuinka hyödylliseksi koit TimeGPT-ennusteen Navita-ennusteen päivittämisessä?



9. Miten TimeGPT-ennusteet vaikuttivat ennusteen tai resurssien suunnitteluun?

**5**  
Vastaukset

10. Mitä lisäkoulutusta tai tukea koet tarvitsevasi ennusteraporttiin tai TimeGPT-ennustemalliin liittyen?

3

Vastaukset

11. Onko sinulla muita huomioita tai kehitysehdotuksia ennusteraporttiin tai TimeGPT-ennustemallian liittyen?

2

Vastaukset

12. Suositteletko Power BI ennusteraportin ja TimeGPT-ennusteen hyödyntämistä työkaverillesi? (1=en, 3=ehkä, 5=kyllä)

4.00

Keskimääräinen  
arvio

