



LSTM- SEKÄ ARIMA-MALLIEN HYÖDYNTÄMINEN KYSYNNÄN SEKÄ MYYNNIN ENNUSTAMISESSA

Forecasting Demand and Sales Using LSTM and ARIMA Models

Lappeenrannan–Lahden teknillinen yliopisto LUT

Tuotantotalouden kandidaatintyö

2025

Justus Helo

Tarkastaja: Tutkijatohtori Antti Ylä-Kujala

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin insinööritieteiden tiedekunta

Tuotantotalous

Justus Helo

LSTM- sekä ARMA-mallien hyödyntäminen kysynnän sekä myynnin ennustamisessa

Tuotantotalouden kandidaatintyö

2025

40 sivua, 8 kuvaa ja 5 taulukkoa

Tarkastaja: Tutkijatohtori Antti Ylä-Kujala

Avainsanat: Koneoppiminen, Syväoppiminen, Ennustemallinnus, Data-analytiikka, Aikasarja, ARIMA, LSTM

Keywords: Machine Learning, Deep Learning, Forecasting, Data Analytics, Time Series, ARIMA, LSTM

Syväoppimismallien hyödyntäminen on yleistynyt viimeisen kymmenen vuoden aikana. Niitä voidaan käyttää ennustemallintamiseen perinteisten tilastollisten mallien lisäksi. Mallien välillä ei ole itsestään selvää, mikä tuottaisi parhaimmat tulokset. Eikä monimutkaisten syväoppimismallien tuloksien muodostumista ole kovin helppo ymmärtää.

Työssä käytiin läpi ARIMA sekä LSTM-mallin välillä vertailua kysynnän sekä myynnin ennustemallinnuksessa systemaattisena kirjallisuuskatsauksena. Työssä tarkennuttiin vertailemaan ennustetarkkuuksia RMSE, MAPE sekä MAE mittareilla 2019 jälkeen kerätyllä aineistolla. Lisäksi työssä tarkennuttiin mallien tulkittavuuteen sekä siihen muodostuneisiin haasteisiin.

Työssä havaittiin LSTM-mallin tuottaneen enimmäkseen parempia ennustetarkkuuksia, mutta sen tulkittavuus oli haasteellisempi kuin ARIMA-mallin. LSTM-mallin havaittiin myös toimivan paremmin suuremmilla dataseiteillä ja kykenevän toimimaan epästationäärisellä datalla. Vastaavasti ARIMA-mallin havaittiin tarvitsevan datan esikäsitteilyä stationäärisyyttä varten, jotta mallin ennustetarkkuudet saataisiin mahdollisimman korkeaksi.

LYHENNELUETTELO

AI	Tekoäly (Artificial Intelligence)
AIC	Akaiken informaatiokriteeri (Akaike Information Criterion)
ANN	Keinotekoinen neuroverkko (Artificial Neural Network)
ARMA	Autoregressiivinen liukuva keskiarvo (Auto Regressive Moving Average)
ARIMA	Autoregressiivinen integroitu liukuva keskiarvo (Auto Regressive Integrated Moving Average)
BIC	Bayesialainen informaatiokriteeri (Bayesian Information Criterion)
BPTT	Backpropagation-tekniikka (Backpropagation Through Time)
CNN	Konvoluutioneuroverkko (Convolution Neural Network)
DL	Syväoppiminen (Deep Learning)
LSTM	Long Short-Term Memory -malli
MAE	Keskimääräinen absoluuttinen virhe (Mean Absolute Error)
MAPE	Keskimääräinen absoluuttinen prosentuaalinen virhe (Mean Absolute Percentage Error)
ME	Keskivirhe (Mean Error)
ML	Koneoppiminen (Machine Learning)
RMSE	Keskivirheen neliöjuuri (Root Mean Square Error)
RNN	Toistuva neuroverkko (Recurrent Neural Network)
SARIMA	Kausittainen Autoregressiivinen integroitu liukuva keskiarvo (Seasonal Auto Regressive Integrated Moving Average)

Sisällysluettelo

Tiivistelmä

Lyhenneluettelo

1	Johdanto	5
1.1	Tausta ja motivaatio	5
1.2	Tutkimuskysymykset sekä tavoitteet	6
1.3	Tutkimusmenetelmät sekä rakenne	7
2	Ennustemallinnus aikasarjadataalle	9
2.1	Ennustemallinnuksen tausta, perusteoria ja tavoitteet	9
2.2	Aikasarjadataan ominaisuudet	10
2.3	Tilastollisten menetelmien sekä koneoppimismallien eroavaisuudet	11
2.4	Ennustemallien arviointimenetelmät	12
3	Tilastolliset menetelmät ennustemallinnuksessa	15
3.1	ARMA-mallien toiminta	15
3.2	Datan vaatimukset sekä esikäsittely	17
3.3	Mallin hyödyt ja rajoitteet	17
4	Kone- ja syväoppimismallit ennustamisessa	18
4.1	Kone- ja syväoppimismallien eroavaisuudet	18
4.2	LSTM-mallin toiminta	20
4.3	Tarvittavan datan määrä ja käsittely	22
4.4	Mallin hyödyt ja rajoitteet	22
5	ARMA- sekä LSTM-mallien suorituskykyjen vertailu	23
5.1	Aineistoon valitut julkaisut	23
5.2	Ennustetarkkuuden vertailu eri aineistoilla	25
5.3	Tuloksien vertailukelpoisuuden haasteet	27
6	Johtopäätökset	29
	Lähteet	34

1 Johdanto

Kokonaisvaltainen ymmärrys eri ennustemallien vahvuuksista ja heikkouksista on tärkeää jokaiselle aiheen kanssa työskenteleville. Ilman relevanttia osaamista tulokset voivat olla liian optimistisia tai johtaa muutoin erehdyksiin päätöksenteossa. Koneoppimismallit saattavat olla vaikeasti tulkittavissa varsinkin, kun niiden data sekä mallit monimutkaistuvat, mikä voi johtaa niiden hyödyn heikkenemiseen päätöksenteon tukena. (Linardatos et al. 2020; Nandi & Pal 2022) Näin on muodostunut uusi haaste data-analytiikan alalla: kuinka hyvin uudet koneoppimismallit pärjäävät perinteisille tilastollisille menetelmille?

1.1 Tausta ja motivaatio

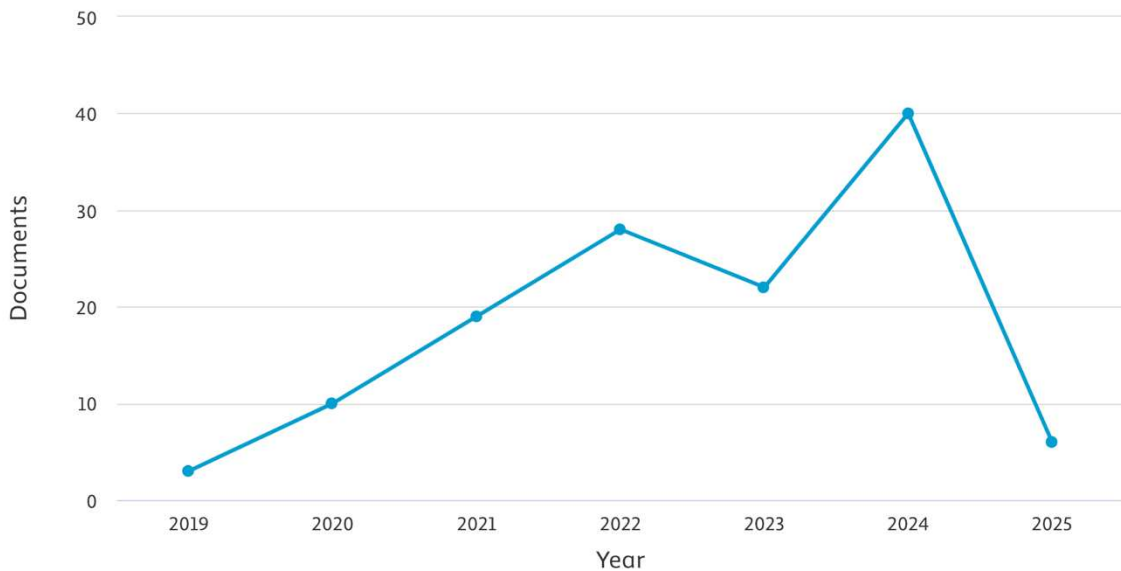
World Economic Forumin viimeisimmän *The Future of Jobs Report 2025* -raportin mukaan tekoäly (AI) sekä koneoppimisen (ML) parissa olevat työt kasvavat nettona yli 80 % välillä 2025–2030 (World Economic Forum 2025). Alan lisääntyvää tarvetta tukee koneoppimismarkkinoiden kasvu sekä taloudellisen epävarmuus EU:ssa (Euroopan komissio 2024; Statista 2024). Oleellinen osa koneoppimista on tuottaa ennusteita halutusta ilmiöstä. Ennen koneoppimista on kuitenkin hyödynnetty tilastollisia menetelmiä ennustemalleissa ekonometriassa, osakekurssien parissa sekä liiketoiminnan päätöksenteon tukena (Gao et al. 2024; Tsay 2020).

Ennustemallinuksia tehdään siis enenevässä määrin, sillä niillä pyritään vähentämään epävarmuutta tulevasta. Niiden tulisi olla tarkkoja, tehokkaita sekä tuottaa haluttuja lopputuloksia. Samalla niiden tulisi olla yksinkertaisesti tulkittavissa (Nandi & Pal 2022). Vastaavaa tarkkuuden käsittelyä on pohdittu jo ennen 1980-lukua, kun ARMA-mallit olivat olleet jo lähes 30 vuotta käytössä (Makridakis et al. 1979). Aihe ei ole siis ainoastaan ajankohtainen vaan myös historiallisesti merkittävä.

Tilastollisilla menetelmillä sekä koneoppimismalleilla pyritään parantamaan ennustetarkkuutta haluttuun ilmiöön, joka tässä työssä rajattu käsittelemään kysyntää sekä myyntiä (Gao et al. 2024; Tsay 2000). Kyseiset aiheet valittiin rajauksen sisälle, sillä ne molemmat ovat yritykselle liikevaihtoa lisääviä tekijöitä. Vaikka Seaman (2018) määrittelee myynnin sekä tuotteen kysynnän eri ennusteiksi, molemmat ovat mukana myynnin sekä kysynnän

ennustamisessa tässä työssä. Vuoden 2018 jälkeen on julkaistu merkittävä määrä vertailevia tutkimuksia ARMA-mallien sekä LSTM-syväoppimismallien välillä, mutta rajaukseen lisättäessä kysynnän sekä myynnin ennustaminen julkaisujen määrä vähenee selkeästi (Kuva 1).

Documents by year



Kuva 1: Aihealueen ARIMA/SARIMA, LSTM, ennustaminen sekä kysyntä julkaisut Scopus viitetietokannasta (Scopus 2025)

Liiketoiminnan sekä finanssialan kohdalla on havaittu haasteita yleistettävyydessä sekä tulkittavuudessa (Gao et al. 2024). Jotta malleja voitaisiin hyödyntää käytännön tasolla, niiden toimintaperiaatteiden tulisi olla tulkittavissa. Syväoppimismallien suorituskyvyn vertautuminen perinteisiin tilastollisiin menetelmiin muuttuu laskentatehon kehityksen myötä, joka muodostaa tutkimusaukon tälle tutkimukselle (Dally et al. 2021). Lisäksi tämä tutkimus pyrkii paikkaamaan myynnin sekä kysynnän ennustemallien tulkittavuuden tutkimusaukkoa. Koska aihe kehittyy jatkuvasti, tämä tutkimus hyödyntää vertailevassa aineistossaan viimeisen kymmenen vuoden aikana julkaistuja tutkimuksia.

1.2 Tutkimuskysymykset sekä tavoitteet

Tämä tutkimus keskittyy tilastollisten menetelmien sekä koneoppimismallien ennustemallien vertailemiseen kysynnän sekä myynnin ennustamisessa. Tarkasteltavat mallit ovat

perinteisiä tilastollisia ARMA-malleja sekä LSTM-syväoppimismalleja. Tutkimuksen pääkysymys on seuraava:

1. Miten LSTM- vertautuu perinteisempiin ARMA-malleihin ennustemallinnuksessa?

Tutkimuskysymyksen vastaavat seuraavat alatutkimuskysymykset:

1.1 Parantuuko ennustetarkkuus LSTM-mallilla?

1.2 Miten mallien tulkittavuudet eroavat?

Ensimmäinen kysymys tarkastelee yleisellä tasolla mallien eroavaisuuksia, kun halutaan ennustaa tuloksia, jotka pohjautuvat aikasarjadataan. Valittujen mallien sekä niihin kohdistuvien suorituskykymittarien avulla vastataan toiseen päätutkimuskysymykseen. Kolmannen tutkimuskysymyksen kohdalla pyritään tuomaan malleja käytännöntasolle perehtymällä niiden tulkittavuuksien eroavaisuuksiin sekä niiden mahdollisiin haasteisiin. Näin tutkimuksella saavutetaan kokonaiskuva valittujen menetelmien hyödyistä sekä haitoista.

1.3 Tutkimusmenetelmät sekä rakenne

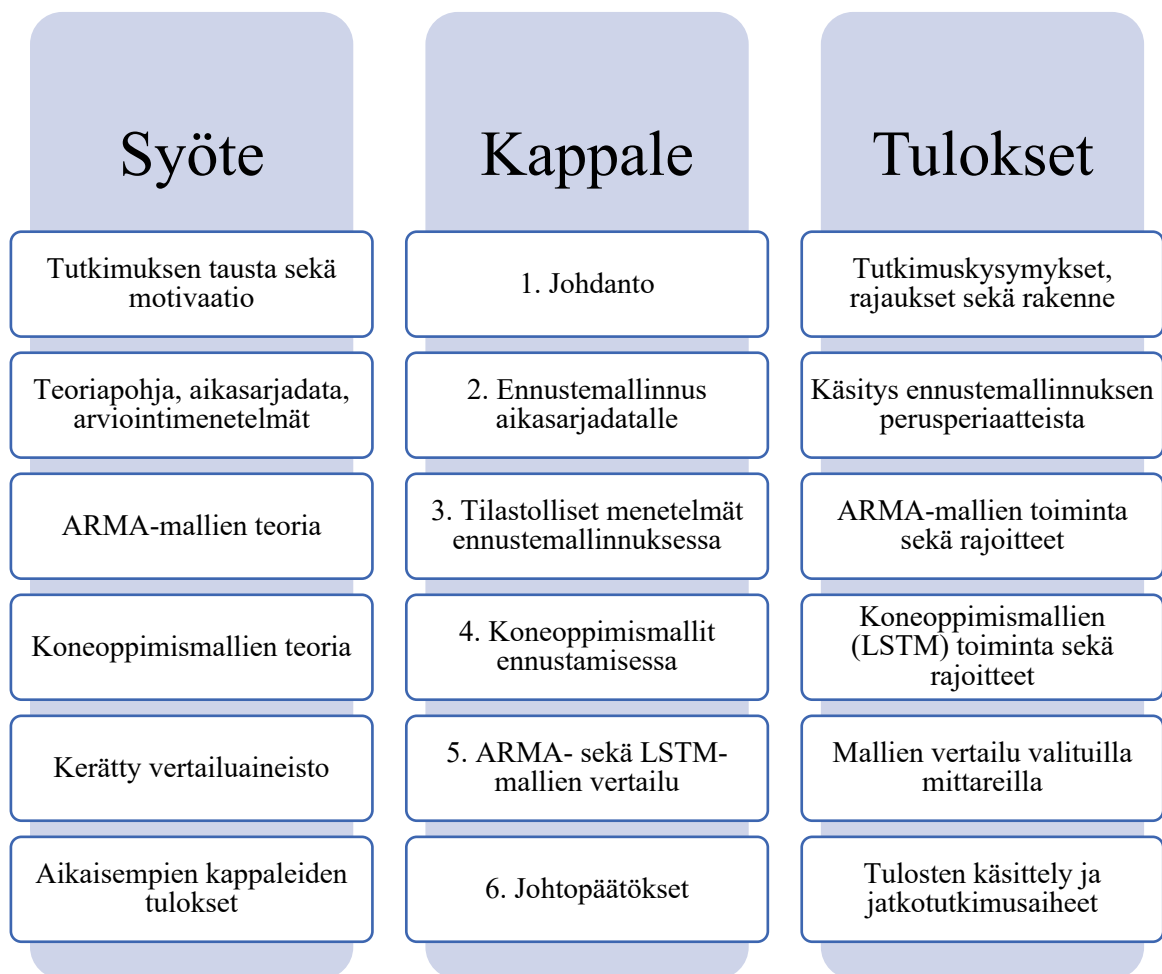
Tämä kandidaatintyö toteutetaan systemaattisena kirjallisuuskatsauksena, jonka aineisto on kerätty 2019 jälkeen julkaistujen tutkimusten perusteella. Työn perusteoriaa kattavat osiot sisältävät myös aiempaa kirjallisuutta. Aineisto on kerätty LUT Primo:n, Scopus- sekä IEEE-viitetietokannan avulla (Taulukko 1). Malleiksi rajattiin tilastollisista menetelmistä ARMA- ja syväoppimismenetelmistä LSTM-mallit. Molemmat mallit ovat parametrisia ja soveltuvat aikasarjadatan ennustamiseen.

Taulukko 1: Käytetyt hakukoneet sekä -sanat

Hakukoneet sekä tietokannat	Hakusanat
LUT Primo	ARMA/ARIMA/SARIMA/LSTM
Scopus	Time Series/Forecast/
IEEE	Sales/Customer Demand
	Machine Learning

Tuloksia vertaillaan seuraavilla suorituskykymittareilla: RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error) sekä MAE (Mean Absolute Error). Tutkimukseen valikoitiin julkaisut, jotka ovat vertaisarvioituja ja ne sisältävät vähintään molemmat valikoidut mallit. Näin tutkimuksen tuloksille saadaan paremmin vertailtua mallien suoriutumiskykyä, vaikka tutkimusten välillä on useita eri datasettejä. Julkaisut rajattiin käsittelemään kysyntää sekä myynnin ennustamista.

Työn rakenne koostuu kolmesta osiosta (Kuva 2). Ensiksi käsitellään ennustemallinnuksen perusteoriaa sekä siinä hyödynnettävän datan ominaisuuksia. Seuraavaksi tutkimuksessa käsitellään vertailtavien mallien toimintaa ja rajoitteita, jonka jälkeen niitä vertaillaan keskenään. Lopuksi työ sisältää tuloksia yhteen vetävän osuuden, joka sisältää jatkotutkimusaiheiden käsittelyä sekä mallien käytännöntason tarkastelua päätöksenteon tukena.



Kuva 2: Tutkimuksen rakenne

2 Ennustemallinnus aikasarjadataalle

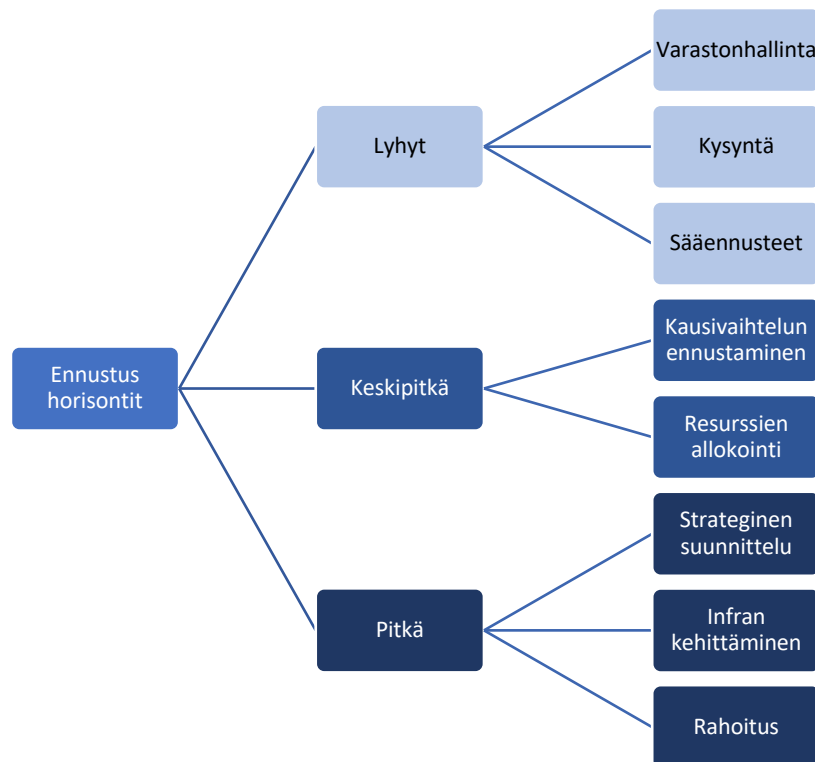
Seuraavat kappaleet käsittelevät ennustemallinnuksen perusteoriaa, menetelmiä eri aloilla sekä eri menetelmiä tulosten arviointia varten. Lisäksi kappaleissa kuvataan aikasarjadataan erityispiirteitä, ja miten se eroaa muista aineistotyypeistä.

2.1 Ennustemallinnuksen tausta, perusteoria ja tavoitteet

Luonnontieteiden alainen ennustaminen sekä tilastollinen analyysi sai alkunsa jo ennen 1900-lukua (Gilliland et al. 2021; Tsay 2000). Ala käy läpi parhaillaan murrosta laskennallisen tehon kehityksen myötä. Yritysten välinen kilpailu lisää tarvetta ennustaa kysyntää ja siitä aiheutuvaa myyntiä. Niiden rooli on myös merkittävää terveydenhuollon alalla, jossa ennustettavia kohteita on epidemioiden muodostumisesta, puuttuvien arvojen laskentaan sekä varastojen tasoihin (Jin et al. 2023; Niako et al. 2024; Mbonyinshuti et al. 2024; Picialli et al. 2021).

Jokaista tieteiden alaa yhdistää ennustamisessa käytettyjen mallien lisäksi epävarmuuden käsittelyn tarve (Carvalho et al. 2015; Gilliland et al. 2021; Goh 2024; Leffrang & Müller 2024). Ennusteet eivät pysty olemaan täydellisiä, joten epävarmuuden huomiotta jättäminen voi johtaa virheellisiin päätöksiin. Mahdollisia haasteita voi olla mallien rajoitteissa, satunnaismuutoksissa sekä ulkoisesti vaikuttavissa tekijöissä. Mitä paremmin epävarmuus on tunnistettu ja mitattavissa, sitä helpommin se on kohdattavissa ennusteiden käsittelyssä (Gilliland et al. 2021).

Ennustehorisontit voidaan luokitella piste- sekä monivaihe-ennusteisiin. Piste-ennusteissa kohdistetaan horisontti yksittäiseen ajankohtaan ja monivaihe-ennusteissa ennustetaan useampia ajankohtia halutulle aikavälille. Horisontit voidaan myös jaotella niiden pituuksien mukaan lyhyisiin, keskipituisiin sekä pitkiin ennusteisiin (Kuva 3). Lyhyellä horisontilla voidaan ennustaa seuraavan päivän kysyntää ja siten kohdistaa resursseja tuotantoon oikeille tuotteille. Keskipitkällä horisontilla voidaan paremmin suunnitella resurssien tarvetta tai ennakoida kausivaihtelusta aiheutuvia muutoksia. Vastaavasti pitkän aikavälin ennusteet voivat kohdistua enemmän strategiaan tai rahoitukseen.



Kuva 3: Esimerkkejä ennustehorisonteista ja käyttökohteista

2.2 Aikasarjadatan ominaisuudet

Aikasarjadataa käytetään ennustemallinnuksessa, kun tavoitteena on mallintaa halutun ilmiön kehittymistä aikaisempien havaintojen perusteella. Se voi olla päivittäistä, kuukausittaista, vuosittaista tai muulla aikajänteellä kerättyä dataa. Aikasarjadata voidaan luonnehtia koostuvan deterministisestä osasta, joka kuvaa säännönmukaista trendiä ja mahdollista kausivaihtelua, sekä stokastisesta osasta, joka edustaa ennakoimatonta ja satunnaista vaihtelua. Se eroaa poikkileikkausaineistosta siten, että siinä on useita ajanhetkiä yksittäisen tai muutoin lyhyen osion sijaan. Aikasarjadata voi sisältää yksittäisen muuttujan tai useita muuttujia, jotka voivat olla toisistaan riippuvia. (Lütkepohl & Krätzig 2004)

Yleensä tilastolliset menetelmät olettavat aikasarjadatan olevan stationäärinen. Tämä tarkoittaa, että sen keskiarvo, varianssi sekä kovarianssi pysyvät vakioina aikasarjan läpi, eli sarjan tilastolliset ominaisuudet eivät ole riippuvaisia tarkasteluhetkestä. Stationäärisyys ei tarkoita, että peräkkäiset arvot olisivat vakioita, vaan näiden tilastollisten ominaisuuksien muuttumattomuutta pidemmällä aikavälillä. Lyhyellä aikavälillä voi siis olla havaittavissa varianssia. Mikäli aikasarja ei ole stationäärinen, se voi aiheutua trendistä tai

kausivaihtelusta. Tämä voi johtaa epäluotettaviin tuloksiin mallin ennustaessa tietyn lyhyemmän ajanjakson erityispiirrettä eteenpäin. Kausivaihtelu ei suoraan ole este ennustamiselle, jos se on säännöllistä ja voidaan mallintaa erikseen. Aikasarjan stationäärisyys voidaan tarkastaa autokorrelaatio- sekä osittaisella autokorrelaatiofunktiolla, jotka mittaavat aiempien arvojen sekä tulevien arvojen välistä korrelaatiota. Stationäärisyys voidaan myös testata tilastollisilla menetelmillä, kuten Dickey-Fuller tai Augmented Dickey-Fuller. Aineiston muokkaaminen stationääriseksi tapahtuu poistamalla trendi differensoinnilla, jossa aikaisempi arvo vähennetään nykyisestä arvosta. Mikäli havaitaan kausivaihtelua, se saadaan muokattua stationääriseksi differensoimalla nykyinen arvo esimerkiksi vuoden takaisesta arvosta. (Cipra 2020; Lütkepohl & Krätzig 2004; Mignon 2024)

Aikasarjadataa hyödyntäviin malleihin voidaan lisätä ulkoisia tekijöitä eli eksogeenisiä muuttujia tehostamaan mallien ennustustarkkuutta. Eksogeenisiä muuttujia voi olla muun muassa sääolosuhteet tai makrotaloudelliset muuttujat kuten korkotasot tai inflaatio (Yadav 2024; Yan et al. 2023). Tällöin on oleellista varmistaa, että eksogeeniset muuttujat eivät ole korreloituneita keskenään sekä aikasarjadatan kanssa. Muutoin malleissa voi ilmentyä multikollinearisuutta, joka voi lisätä varianssia sekä virheitä tilastollisissa testeissä (Lütkepohl & Krätzig 2004). Eksogeenisten muuttujien tulee olla myös ennustettavissa ja kohdistettu samalle aikajänteelle, jotta malli säilyy luotettavana.

2.3 Tilastollisten menetelmien sekä koneoppimismallien eroavaisuudet

Tilastolliset menetelmät ja koneoppimismallit ovat yleisesti käytettyjä lähestymistapoja ennusteiden mallintamiseen. Tilastolliset menetelmät ja mallit pohjautuvat matemaattisiin kaavoihin sekä tilastollisiin oletuksiin kuten datan stationäärisyys, lineaarisuuteen sekä normaali jakaumaan (Lütkepohl & Krätzig 2004). Myös koneoppimismallit hyödyntävät matemaattisia kaavoja, mutta ne eivät ole aina yhtä tarkasti ennalta määrättyjä tai vaadi samoja tilastollisia oletuksia.

Ne voidaan myös luokitella parametrisiin sekä ei parametrisiin malleihin (Taulukko 2). Parametristen mallien matemaattiset kaavat ovat ennalta määriteltäviä. Ei parametriset mallit oppivat riippuvuudet sekä matemaattiset mallit datasta. Koneoppimismalli ei kuitenkaan ole välttämättä täysin parametrinen vaan voi sisältää ei parametrisiä osia. Niissä parametrit eivät aina ole ennalta määrättyjä, vaan ne voivat määräytyä mallin oppimisen kautta (Khadse et

al. 2020). Parametriset mallit täten myös vaativat vähemmän laskennallista tehoa verrattuna ei parametrisiin (Gautam & Singh 2020; Seaman 2018). Koneoppimismallit ovat herkempiä mallien ylisovittamiselle, joka voi johtaa optimistisiin tuloksiin koulutusdatalla ilman samoja tuloksia testidatalla (Fildes et al. 2022).

Taulukko 2: Esimerkkejä parametrisista sekä ei parametrisista malleista aikasarjadatan ennustamiseen

Menetelmä	Parametriset mallit	Ei parametriset mallit
Tilastolliset mallit	ARMA-mallit Vektori autoregressio (VAR) Lineaarinen ja logistinen regressio	LOESS Kernel regressio
Kone- ja syväoppimismallit	Tukivektorikoneet (SVM) LSTM/GRU (Ei täysin parametrisinen)	Decision Tree (DT)/Random Forest (RF) Gradient Boost -mallit (GB) K-Nearest Neighbors (KNN)

Tilastolliset menetelmät ja koneoppimismallit eroavat myös niiden laskentatehon vaatimuksilta. Tilastolliset menetelmät vaativat vähemmän laskennallista tehoa vähäisempien parametrien takia. Niissä on valmiiksi rajattu määrä parametreja, joiden avulla malli optimoidaan.

2.4 Ennustemallien arviointimenetelmät

Ennusteiden arviointi on keskeinen osa mallien tarkastelua sekä suorituskyvyn mittaamista. Ilman riittävää arviointia malli voi tuottaa harhaan johtavia tuloksia, jotka voivat johtaa virheellisiin päätöksiin, vaikka malli olisikin tarkkuusmetriikaltaan hyvä (Nandi & Pal 2022). Oleellista on pyrkiä valitsemaan tapaukseen sopivat arviointimenetelmät mallinnuksen eri vaiheisiin. Arvioinnin onkin siis hyvä alkaa jo datan käsittelystä.

Datan tulisi olla puhdistettu merkittävistä outlier-arvoista, sillä muutoin malli voi oppia ne ja vääristää tuloksia (Cissoko et al. 2025). Dataa syöttäessä mallille tulisi ottaa huomioon ristivalidaatio, sillä se lisää mallin ennusteiden tarkkuutta (Makridakis et al. 2022). Mallin

tarkkuutta arvioidessa tulisi valikoida tapaukselle sopivat mittarit. Väärin valitut mittarit voivat johtaa mallin ylisovittamiseen tai olla käyttökelvottomia datan luonteen takia. Esimerkiksi keskimääräinen absoluuttinen prosentuaalinen virhe (MAPE) ei ole sopiva mittari datalle, jonka arvot pyörivät nollan lähellä (Hewamalage et al. 2023). Sen arvot kasvavat liian suuriksi kaavansa (1) takia, kun ennustettu arvo Y_t ja todellinen arvo y_t ovat alle 1, ja mahdolliset negatiiviset arvot eivät anna luontevasti tulkittavia positiivisia tuloksia (Hewamalage et al. 2023).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - Y_t}{y_t} \right| * 100 \% \quad (1),$$

jossa ennustettu arvo on Y_t ja todellinen arvo y_t .

Ristivalidaation on tärkeätä, jotta mallit olisivat paremmin yleistettävissä. Sen peruseriaate on jakaa data koulutus- sekä testiosioon (Liu & Zhou 2024). Yksi menetelmä on Holdout, jossa data jaetaan yhteen koulutus- ja testiosaan. Ristivalidaatioon yleisempi menetelmä on K-fold, jossa data jaetaan valittuun määrään koulutus- ja testiosia (Kohavi 1995; Bergmeir et al. 2018). K-fold ei ole kuitenkaan sopiva aikasarjadataalle, joka on autokorreloitunutta (Vamsikrishna & Gijou 2024). K-fold jakaa koulutus- ja testiosiot satunnaisesti, jolloin datasta katoaa peräkkäisten havaintojen korrelaatio. Bergmeir et al. (2018) mukaan K-fold voisi olla mahdollinen metodi, mikäli data on puhtaasti autokorreloitunut ja sen jäännösarvot eivät ole korreloituneita keskenään. Lisäksi ristivalidaation metodia valittaessa tulee ottaa huomioon aikasarjan stationäärisyys, sillä epästationääriselle datalle voi olla parempi valita eri metodi (Cerqueira et al. 2020). Aikasarjadataalle käytettyjä ristivalidaation metodeja on muun muassa liukuva ennuste (rolling forecast) sekä kasvava ikkuna (walk forward) (Mehrab et al. 2021; Wang et al. 2021). Reaalimaailman tapauksissa data ei kuitenkaan usein ole stationääristä vaan sisältää kausivaihtelua sekä muita varianssin vaihtelua aiheuttavia tekijöitä.

Ennustetarkkuuden tai -virheen arvioinnissa ei ole yhtä selkeätä mielipidettä parhaasta mittarista (Makridakis et al. 2022). Yleisesti käytettyjä mittareita MAPE:n lisäksi ovat keskivirhe (MSE) (2), keskivirheen neliöjuuri (RMSE) (3) sekä keskimääräinen absoluuttinen virhe (MAE) (4), millä on jokaisella omat vahvuutensa ennustevirheiden arvioinnissa (Taulukko 3). (Koutsandreas et al. 2021)

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - Y_t)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - Y_t)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - Y_t| \quad (4),$$

joissa ennustettu arvo on Y_t ja todellinen arvo y_t .

Taulukko 3: Yleisiä ennustevirheiden arviointimenetelmiä

Arviointimenetelmä	Kaava	Menetelmien etuja sekä rajoitteita
MAPE	$\frac{1}{n} \sum_{t=1}^n \left \frac{y_t - Y_t}{y_t} \right * 100 \%$	Helppo tulkita virheen suuruutta suhteessa todellisiin arvoihin. Huono pienillä arvoilla, sillä voi johtaa epätarkkoihin tuloksiin.
MAE	$\frac{1}{n} \sum_{t=1}^n y_t - Y_t $	Kertoo suoraan keskimääräisen poikkeaman todellisista arvoista. Ei yhtä herkkä suurille poikkeamille, joten RMSE voi olla parempi tarkemmissa tapauksissa.
MSE	$\frac{1}{n} \sum_{t=1}^n (y_t - Y_t)^2$	Korostaa suuria poikkeamia neliöinnin takia enemmän kuin RMSE. Hyvä mallien optimoinnissa herkkyyden vuoksi. Ei yhtä intuitiivinen, kun RMSE yksiköiden neliöinnin takia.
RMSE	$\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - Y_t)^2}$	Hyödyllinen kriittisemmissä ennusteissa herkkyytensä vuoksi. Helpompi tulkita, mutta korostaa vähemmän virheitä verrattuna MSE:hen.

3 Tilastolliset menetelmät ennustemallinnuksessa

Aikasarjadataan ennustaminen tilastollisilla menetelmillä sisältää lukuisia eri malleja, kuten ARIMA, vektori-autoregressio (VAR) sekä eksponentiaalinen tasoitus. Mallien periaatteet ovat hieman toisistaan poikkeavia, vaikka niitä voidaan soveltaa samoihin käyttökohteisiin. Seuraavat kappaleet käsittelevät autoregressiivisen liukuvan keskiarvon mallin eri muotoja sekä käyttöä, mikä on jatkossa lyhenteellä ARMA.

3.1 ARMA-mallien toiminta

ARMA-mallit koostuvat vähintään autoregressiivisestä osasta (AR) sekä liukuvan keskiarvon osasta (MA), joiden avulla voidaan ennustaa ja mallintaa aikasarjadataa. Autoregressiivinen komponentti (AR) viittaa mallin kykyyn ottaa huomioon aikaisemman arvon vaikutuksen nykyiseen arvoon. Liukuvan keskiarvon komponentti (MA) puolestaan mallintaa aikaisempien residuaalien vaikutusta nykyiseen arvoon liukuvasti. Eri mallit sisältävät muutoksia matemaattisiin kaavoihin lisäten niihin esimerkiksi trendin poiston, eksogeenisen muuttujan tai kausivaihtelun poiston. Yksinkertaisin malleista on ARMA (5), joka ei sisällä differensointia tai kausivaihtelun poistoa.

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j y_{t-j} + \epsilon_t \quad (5),$$

jossa $\phi_1 \dots \phi_i$ ovat autoregressiiviset (AR) painot, $\theta_1 \dots \theta_j$ ovat liukuvan keskiarvon (MA) painot, p on AR-osan viiveiden määrä, q on MA-osan viiveiden määrä.

ARMA-mallit tulivat tunnetuksi Box-Jenkins-menetelmästä, joka pyrkii rakentamaan mallin systemaattisen prosessin kautta. Se voidaan jakaa kolmeen päävaiheeseen: mallin valinta, parametrien määrittely sekä mallin arviointi. Mallin valinta toteutuu tarkastelemalla aikasarjan stationäärisuys. Mikäli trendiä tai kausivaihtelua on havaittavissa, muutetaan malli siihen sopivaksi (Taulukko 4). Sopivat viiveet p ja q määritetään autokorrelaatiofunktion sekä osittaisautokorrelaatiofunktion avulla. Parametrit määritellään esimerkiksi suurimman uskottavuuden estimaatilla. Ja lopulta mallista tarkastetaan jännösten normaalijakautuneisuus sekä autokorrelaatio. Niiden tarkastaminen on tärkeätä tulosten luotettavuuden validoimiseksi. (Box & Jenkins 1976)

Taulukko 4: Yleisimpiä ARMA-malleja

Malli	Merkintä	Selite
Autoregressiivinen malli (AR)	AR (p)	Nykyinen arvo riippuu menneistä arvoista.
Liikkuva keskiarvo (MA)	MA (q)	Nykyinen arvo riippuu menneistä residuaaleista
Autoregressiivinen liikkuva keskiarvo (ARMA)	ARMA (p, q)	Yhdistää AR ja MA samaan malliin
Autoregressiivinen integroitu liikkuva keskiarvo (ARIMA)	ARIMA (p, d, q)	Lisää trendin poistavan differensoinnin (d) ARMA malliin
Kausittainen ARIMA (SARIMA)	SARIMA (p, d, q) (P, D, Q, s)	Lisää kausivaihtelun poiston malliin.

Suotuisampien parametrien valinta vie enemmän aikaa siirtyessä ARMA-mallista monimutkaisempiin ARIMA- sekä SARIMA-malleihin. Ennustetarkkuuden sekä mallin sovituksen välistä yhteyttä tulisi tarkastella kvantitatiivisesti, jotta valittu malli olisi perustellusti paras. Yleisesti käytettyjä menetelmiä ovat Akaiken informaatiokriteeri (AIC) sekä Bayesialainen informaatiokriteeri (BIC) (Kuha 2004). Niiden avulla voidaan tarkastella mallin tarkkuuden sekä parametrien tasapainoa. Mikäli malli valitaan BIC avulla eikä AIC, malli voi kärsiä ennustetarkkuudessaan. AIC arvon perusteella valittu malli toisaalta voi olla ylisovitettu. AIC- sekä BIC-arvoja hyödyntäessä paras malli valitaan pienimmän arvon mukaan. AIC (6) ja BIC (7) poikkeavat vain hieman kaavaltansa, sillä BIC ottaa huomioon mallin aineiston havaintojen määrän.

$$AIC = -2 \ln(L) + 2k \quad (6)$$

$$BIC = -2 \ln(L) + k \ln(n) \quad (7),$$

joissa L on log-uskottavuusfunktio, k on parametrien määrä ja n on havaintojen määrä.

3.2 Datan vaatimukset sekä esikäsittely

ARMA-mallia käyttäessä oletuksena on datan stationäärisyys (Lütkepohl & Krätzig 2004). ARMA-mallille on olemassa vaihtoehtoisia malleja, joilla aineisto voidaan muokata mallin sisällä stationääriseksi. ARIMA-malli sisältää trendin poistavan differensoinnin ja SARIMA- ottaa huomioon kausivaihtelun poistamisen valitulla viiveellä. Datan esikäsittelyssä tulisi ottaa huomioon merkittävien poikkeamien sekä puuttuvien arvojen käsittely. Puuttuvat arvot tulisi täyttää, jotta aikasarja pysyisi ehyenä. (Bianco et al. 2001; Laome et al. 2021; Niako et al. 2024)

ARMA-mallit olettavat, että jäännöstermien varianssi pysyy vakiona. Jos mallissa on havaittavissa heteroskedastisuutta, aineistolle voidaan tehdä logaritmuunnos tai Box-Cox-muunnos (Cipra 2020). Esikäsittelyssä tulisi kuitenkin huomioida, että liikaa alkuperäisestä tiedosta ei poistu. Muutoin malli ei enää tuota välttämättä hyödyllisiä tuloksia.

3.3 Mallin hyödyt ja rajoitteet

ARMA-mallien merkittävä etu on niiden yksinkertaisuus sekä tulkittavuus. Ne tarvitsevat vähimmillään noin 50 havaintoa, joka on merkittävästi vähemmän verrattuna kone- ja syväoppimismalleihin (Box & Tiao 1975; Master's in Data Science n.d.). Malli koostuu selkeistä AR- ja MA-osista, joiden viivearvoilla p ja q saadaan määriteltyä huomioon otettujen aikaisempien havaintojen määrä. Sen laajennuksilla, kuten ARIMA ja SARIMA, saadaan vähäisillä parametriläyksillä otettua huomioon kausivaihtelua sekä trendiä. Mallit ovat matemaattisesti melko yksinkertaisia ja tämä tekee sen tuloksista helposti tulkittavia. Niiden koulutus pohjautuu perinteisiin uskottavuusfunktioihin, eivätkä ne tarvitse laskennallisesti vaativia syväoppimisalgoritmeja parametrien määrittelyyn.

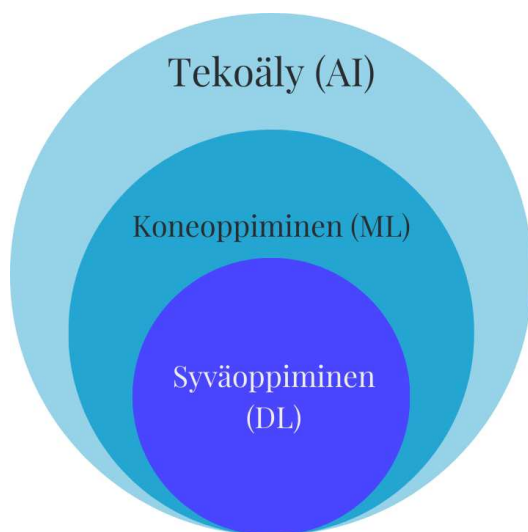
ARMA-mallin käytössä on myös rajoitteita, jotka voivat johtaa ylisovittamiseen. Mikäli parametrit p ja q määritetään liian suuriksi, malli oppii liikaa datasta. Vastaavasti liian suuri differensointi tai virheellinen kausivaihtelun komponenttien valinta voi heikentää mallin tarkkuutta sekä yleistettävyyttä. Näissä tapauksissa malli voi alkaa ylisovittamaan satunnaisvaihtelua, jota ei oikeasti ole. Tämän takia ARMA-malleja käyttäessä tulisi tarkastaa jäännösarvojen korrelaatio. Mikäli korrelaatiota on havaittavissa, malli voi olla ylisovittunut ja oppinut liian tarkasti satunnaista vaihtelua, jolloin sen ennustetarkkuus heikkenee.

4 Kone- ja syväoppimismallit ennustamisessa

Seuraavissa kappaleissa käsitellään yleisiä kone- ja syväoppimismalleja aikasarjadataalle ennustamiseen. Kappaleet tarkentuvat Long Short-Term Memory -mallin (LSTM) käsittelyyn, sillä se on valikoitu tähän työhön vertailtavaksi ARMA-mallien kanssa. Kappaleissa käsitellään datan esikäsittelyä sekä LSTM-mallin hyötyjä sekä rajoitteita aikasarjadataan ennustamismalleissa.

4.1 Kone- ja syväoppimismallien eroavaisuudet

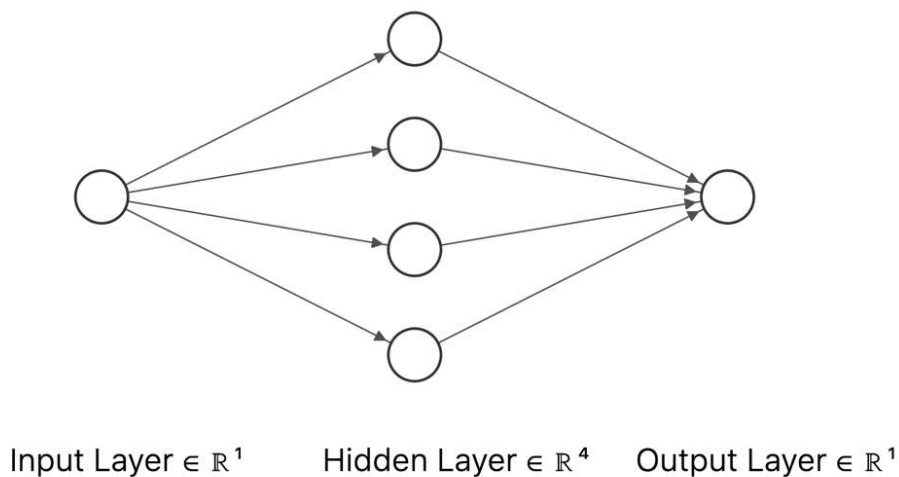
Koneoppiminen voidaan luokitella tekoälyn osajoukoksi, jonka sisälle kuuluu syväoppiminen (Kuva 4). Kone- ja syväoppimista yhdistää datan hyödyntäminen algoritmien kouluttamiseksi. Niiden välillä on kuitenkin eroavaisuuksia toimintaperiaatteissa, tulkittavuudessa sekä datan tarpeissa ja laskentatehon vaatimuksissa. Kone- ja syväoppimismallien kouluttaminen voi olla valvottua, valvomaton tai vahvistavaa oppimista. Valvotussa oppimisessa malli tietää datasta selitettävän muuttujan etukäteen ja se pyrkii mallintamaan annetun syötteen välistä yhteyttä siihen. Valvottoman oppiminen puolestaan perustuu datan kategorisointiin. Vahvistava oppiminen on monimutkaisempi järjestelmä, jossa ”agentti” eli järjestelmä oppii kokeilemalla toimintoja ympäristössä. (Auffarth 2021; Jo 2021; Nandi & Pal 2022)



Kuva 4: Tekoälyn, kone- ja syväoppimisen suhde. (Mukaillen Khalkar et al. 2021)

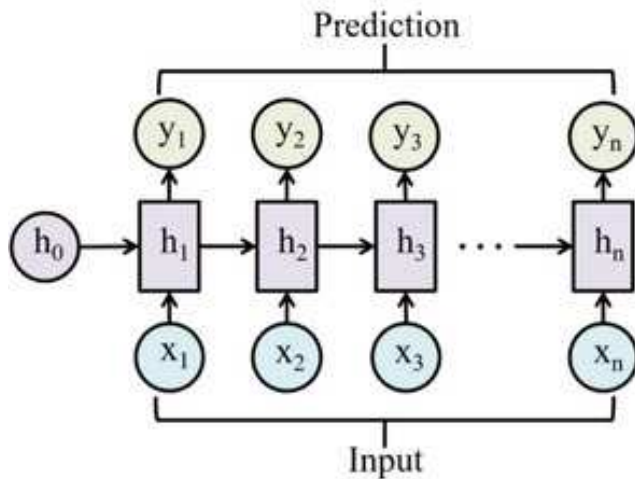
Koneoppimismallit pohjautuvat tilastollisiin menetelmiin sekä matemaattisiin algoritmeihin, jotka oppivat datan avulla tuottamaan haluttuja lopputuloksia. Yleisesti käytettyjä koneoppimismalleja ovat regressiomallit, päätöspuut (decision tree), satunnaismetsät (random forest) sekä k-lähimmän naapurin malli (k-nearest neighbour). Niille riittää pienempi aineisto ja ne tarvitsevat vähemmän laskennallista tehoa verrattuna syväoppimismalleihin.

Syväoppimismallit eroavat koneoppimismalleista hyödyntämällä keinotekoisia neuroverkkoja (ANN). Niiden toiminta perustuu verkkomaiseen rakenteeseen, jossa syötekerroksen (input layer) data muokataan piilokerrosten (hidden layer) läpi ulostulokerrokseen (output layer) (Kuva 5). Syötekerrokseen saapuva data muokkautuu ulostuloksi aktivointifunktioiden sekä bias-termien avulla, jotka sijaitsevat piilokerroksen neuroneissa.



Kuva 5: Yksinkertainen kuvaus neuroverkosta

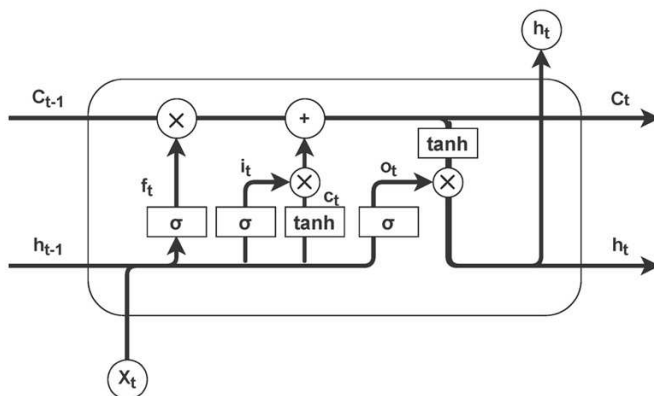
Rekursiiviset neuroverkot (RNN) ovat neuroverkkojen tapaus, joka kykenee säilyttämään aikaisempien havaintojen tietoa muistikerrokseen h_0 (Kuva 6) (Mienye et al. 2024). Muistikerroksen avulla malli oppii datasta ajallisia riippuvuuksia. RNN-mallit kärsivät kuitenkin katoavan tai räjähtävän gradientin ongelmasta (vanishing/exploding gradient problem) (Hochreiter 1998; Mienye et al. 2024). Katoavan gradientin ongelmassa neuroverkon painotukset pienenevät liikaa, joka johtaa mallin kykenemättömyyteen oppia pidemmän aikavälin riippuvuuksia. Räjähtävän gradientin ongelma puolestaan kasvattaa painotukset liian suuriksi, jolloin voi seurata liian suuria muutoksia ja malli ei kykene tuottamaan tarkkoja tuloksia. Hochreiter ja Schmidhuber (1997) kehittivät Long Short-Term Memory-mallin (LSTM) ratkaisemaan kyseisiä ongelmia.



Kuva 6: RNN mallin arkkitehtuuri (Mienye et al. 2024)

4.2 LSTM-mallin toiminta

Long Short-Term Memory (LSTM) on rekursiivisen neuroverkon (RNN) erikoistyyppi, joka kykenee käsittelemään pitkäaikaisia riippuvuuksia datassa. Se eroaa tavallisista RNN-malleista kykenemällä ratkaisemaan katoavan gradientin ongelman (vanishing gradient problem) solurakenteensa takia (Mienye et al. 2024). Lyhytaikainen (hidden state) sekä pitkäaikainen muisti (cell state) ovat LSTM-mallin erityispiirre, jonka avulla aiempien havaintojen tieto säilyy mallin sisällä kolmen portin avulla. Unohtamisportti määrittelee uuden syötteen sekä lyhytaikaisen muistin perusteella paljonko aiempaa tietoa poistetaan. Sitten tulosportti määrittelee ulostuloon menevän tiedon ja lopulta syöteportti päivittää pitkäaikaisen sekä lyhytaikaisen muistin. (Sherstinsky 2020; Mienye et al. 2024) LSTM-mallin arkkitehtuuri on esitetty kuvassa 7.



Kuva 7: LSTM-mallin arkkitehtuuri (Oliveira et al. 2021)

Neuroverkoissa tieto kulkee aktivaatiofunktioiden läpi, sillä ne mahdollistavat tiedon ei-lineaarisen käsittelyn. LSTM-mallin porttien läpi kulkeva tieto päivittyy aktivaatiofunktioiden läpi seuraavasti. Unohtamisportti ensiksi määrittelee paljonko C_{t-1} pitkäaikaisesta muistista säilytetään kaavan (8) mukaisesti.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8),$$

jossa f_t on unohtamisportin aktivaatiofunktio, h_{t-1} on edellisen aika-askeleen lyhytaikainen muisti, x_t on nykyinen syöte, W_f ja b_f ovat painot sekä bias-termit ja σ on sigmoidi-funktio. Sigmoidi-funktio saa arvonsa 0–1 välillä. Sitten syöteportin läpi määritetään uuden tiedon lisäyksen määrä pitkäaikaiseen muistiin kaavojen (9 ja 10) mukaisesti.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10),$$

joissa i_t on syöteportin aktivaatiofunktio, c_t on uusi pitkäaikaisen muistin ehdotettu tieto ja painot ja bias-termit ovat W_i , W_c , b_i ja b_c . Tanh-funktio saa arvonsa -1 ja 1 välillä. Pitkäaikainen muisti päivittyy seuraavasti kaavan (11) mukaisesti.

$$C_t = f_t \odot C_{t-1} + i_t \odot c_t \quad (11),$$

jossa C_t on päivitetty pitkäaikainen muisti ja \odot on Hadamardin tulo. Lopulta tulosportin läpi saadaan haluttu arvo sekä päivitetty lyhytaikainen muisti kaavojen (12 ja 13) mukaisesti.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = o_t \odot \tanh(C_t) \quad (13),$$

joissa o_t on tulosportin aktivaatiofunktio ja h_t on lyhytaikainen muisti sekä ulostulo.

LSTM-mallin suorituskykyyn vaikuttaa sen hyperparametrien määrittely, joka toteutetaan ennen mallin kouluttamista. Mahdollisia hyperparametreja on esimerkiksi piilotilojen määrä (number of hidden units), oppimisnopeus (learning rate) sekä Epoch. Piilotilojen määrä kuvastaa painotermien ulottuvuuksien määrää, joten suurempi luku mahdollistaa monimutkaisempien riippuvuuksien mallintamisen. Oppimisnopeus liittyy painotermien päivittämiseen, joka tapahtuu BPTT-tekniikalla (Backpropagation Through Time). Epoch puolestaan kertoo datan läpikäyntien määrän koulutusvaiheessa.

4.3 Tarvittavan datan määrä ja käsittely

LSTM vaatii merkittävästi enemmän dataa verrattuna ARMA-malleihin. Havaintoja tulisi olla yli tuhat, jotta mallin parametrit saataisiin tarkennettua riittävästi. Liian pienellä datalla malli ei välttämättä tunnista riittävästi vaihtelua. LSTM-malleja käytettäessä tulee tarkastella puuttuvat arvot ja täyttää ne tilanteeseen sopivalla tavalla. Datan ei tarvitse olla stationääristä, kun käytetään LSTM-malleja. Stationääriseksi muokattu data voi kuitenkin nopeuttaa mallin oppimisprosessia, kun trendi tai kausivaihtelu on jo valmiiksi poistettu eikä sitä tarvitse oppia tai unohtaa.

Datan käsittelyssä merkittävä tekijä on ristivalidaatio, jota käsiteltiin tarkemmin kappaleessa 2.4. Sen avulla saadaan paremmin varmistettua, että mallista ei muodostu ylisovitettu. Aikasarjadatan mallintamisessa oleellinen tekijä on tulevien arvojen ennustaminen, joita malli ei ole vielä nähnyt. Mikäli malli on koulutusvaiheessa nähnyt kaiken datan, se voi johtaa epä-säännöllisen vaihtelun liian tarkkaan ennustamiseen. Ristivalidaatiolla on myös merkitys LSTM-mallin hyperparametrien määrittämiseen (Singh et al. 2023).

4.4 Mallin hyödyt ja rajoitteet

LSTM-malleilla on omat hyötynsä aikasarjadatan ennustamisessa ja niillä on mahdollista tuottaa hyvinkin tarkkoja tuloksia. Yksi merkittävimmistä tekijöistä on se, että se ei vaadi stationääristä dataa. LSTM-malli mahdollistaa täten epälineaaristen riippuvuuksien mallintamisen sekä suurien datamäärien käsittelyn. Se tekee LSTM-malleista joustavamman vaihtoehdon verrattuna perinteisiin tilastollisiin menetelmiin, sillä ne pystyvät käsittelemään datassa esiintyvää trendiä sekä kausivaihtelua suoraan mallin sisällä. Tämä on mahdollista sen solurakenteen sekä hyperparametrien avulla.

LSTM-malli kuitenkin vaatii enemmän laskentatehoa verrattuna ARMA-malleihin. Tämä johtuu useampien hyperparametrien määrittämisestä sekä suuremman aineiston tarpeesta. Hyperparametrien optimointi tarvitsee asiantuntemusta sekä omat ristivalidaatiotekniikat, jotta malli ei ylisovitu. (Singh et al. 2023) LSTM-mallit ovat myös vaikeampia tulkita, sillä ne näyttävät mustilta laatikoilta (black box models), joista tiedetään vain lähtötiedot sekä lopputulokset ilman tarkkoja matemaattisia laskuja niiden takana.

5 ARMA- sekä LSTM-mallien suorituskykyjen vertailu

ARMA-mallit ovat helposti käytettäviä benchmark-malleja aikasarjadataan ennustemallien vertailussa, sillä niissä on vähän parametreja ja niiden tulokset ovat helposti tulkittavissa. Seuraavat kappaleet käsittelevät työhön kerättyjen aineistojen tuloksia valituilla mittareilla. Aineisto sisälsi 2019 jälkeen julkaistuja tutkimuksia, jotka olivat vertaisarvioituja ja käsitelivät myynnin tai kysynnän ennustamista. Aineistoista kirjattiin taulukkoon useampi havainto, mikäli niistä löytyi useita ARMA-malleja. Lopulta käsitellään tulosten vertailukelpoisuutta sekä käytettyjen mittarien yleisyyttä aineistojen perusteella.

5.1 Aineistoon valitut julkaisut

Temür et al. (2019) tutkivat julkaisussaan asuntojenmyynnin ennustamista ARIMA, LSTM sekä hybridimallilla. Data oli kerätty 2008–2018 väliltä ja oli kuukauden tarkkuudella 124kk ajalta. Julkaisussa ei mainita tarkempaa trendin tai kausivaihtelun käsittelyä ja datasta käytettiin 70 % mallien koulutukseen ja 30 % testaukseen. ARIMA-mallin parametrit estimoitiin seitsemästä eri mallista muotoon ARIMA (1, 1, 1) tai ARIMA (0, 1, 2), sillä ne tuottivat alimmat MAPE tulokset. ARIMA (1, 1, 1) kuitenkin tuotti paremmat RMSE sekä MSE tulokset. LSTM-malli optimoitiin julkaisussa Epoch-arvojen avulla ja tarkimmat ennusteet tuottivat Epoch (1000). Julkaisu ei sisältänyt ristivalidaation metodeja eikä stationäärisyyden käsittelyä. Se sisälsi myös hybridi mallin (LSTM 1500 Epoch-ARIMA (1,1,1)), joka tuotti malleista tarkimmat tulokset.

Abbamesimehr et al. (2020) käsitelivät useita eri ennustemalleja huonekalujen, teknologia-laitteiden sekä toimistotarvikkeiden myyntidatalla. Data oli kerätty 2014–2017 väliltä ja se koostui lähes 10 000 havainnosta. Tarkkuus muokattiin kuukausitasolle korkean päivittäisen vaihtelun takia. Mittareiksi valikoitiin MAPE, MSE sekä RMSE. 75 % datasta käytettiin koulutukseen ja 25 % testaukseen. ARIMA (6, 0, 0) sekä SARIMA (0, 0, 0)(1, 1, 0, 12) parametrit optimoitiin AIC arvon avulla. Julkaisu sisälsi neljä eri LSTM-mallia, joista yksinkertaisimmassa normaalissa LSTM-mallissa oli yksi piilokerros sekä ulostulokerros. Valittuja hyperparametreja ei kuvattu julkaisussa.

Athiyarath et al. (2020) vertailivat useita eri ennustamismalleja kolmella eri ennustehorisontilla sekä aineistolla. Datan tarkkuustasoa ei mainittu. Tuloksista taulukoitiin nettimyyntiä käsitelleiden havaintojen RMSE arvot. Julkaisu sisälsi RMSE arvon lisäksi selityksasteen (R^2). Datasta mainittiin, että se sisältää ”sekä kausivaihtelua, mutta ei trendiä ja kausivaihtelua”, joten sen voidaan olettaa olevan stationäärinen. Julkaisu ei tarkentanut käytettyjä parametreja eikä ristivalidaation metodeja.

Wang et al. (2021) vertailivat viiden tuotteen kysynnän ennustamista 1.1.2017–31.8.2019 väliltä ja havaintoarvoja oli kokonaisuudessaan 6000. Data muokattiin viikkotasolle ja se ei sisältänyt trendiä tai kausivaihtelua. Julkaisun ennustetavoite oli lyhyellä kolmen viikon horisontilla. Datan esikäsittely sisälsi myös puuttuvien arvojen käsittelyn. Valitut metriikat olivat MAE, MAPE sekä RMSE, mutta julkaisuun kirjoitettu MAE:n kaava oli MSE:n kaava. ARIMA-mallin optimointi toteutettiin AIC ja BIC arvojen tarkastelulla. LSTM-mallin ristivalidaation metodina käytettiin liukuvan ennusteen rakennetta (rolling forecast structure). Sen hyperparametreiksi valikoitiin Epoch (100), otoskoot olivat 95 % koulutus- ja 5 % testidataa ja sen painojen optimoinnin metodiksi valittiin Adam-malli.

Sirisha et al. (2022) käsitelivät myyntidatan ennustamista 1972–2017 väliltä ARIMA, SARIMA ja LSTM-malleja hyödyntäen. Data käsiteltiin kuukausitasolla, joten datapisteitä oli lopulta 548. ARIMA-malli koulutettiin 80 % datalla ja koulutettiin 20 % testidatalla. Datan epästationäärisyys käsiteltiin tekemällä logaritminuunnos. ARIMA:n parametreiksi valittiin (2, 1, 2) tarkastelemalla AR sekä MA osia erikseen ja lopulta yhdistämällä ne. SARIMA:n parametreiksi valittiin (2, 1, 4)(0, 1, 1, 7) ja sillä käytettiin alkuperäistä dataa ilman logaritminuunnosta. LSTM-malli koostui yhdestä näkyvästä kerroksesta, sadasta piilokerroksesta (hidden layer) sekä ulostulokerroksesta. Data MinMax-skaalattiin 0 ja 1 välille LSTM-mallia varten. Sen optimointimalli oli Adam ja paras tulos saatiin Epoch (2), mutta muita koulutukseen merkittäviä hyperparametreja ei mainittu ristivalidaation metodien lisäksi. LSTM-mallin mainittiin olevan tarkempi, mutta kalliimpi toteuttaa ja vaativan enemmän laskennallista tehoa.

Ensafi et al. (2022) tutkivat useita eri ennustemalleja huonekalujen myyntidatalla, joka oli kerätty 2014–2017 väliltä. Se sisälsi 9994 datapistettä ja ennustetarkkuus oli kuukauden keskitasolla. Valitut metriikat olivat MSE, RMSE sekä MAPE. ARIMA:n sekä SARIMA:n parametrit määriteltiin Grid Search-metodilla seuraavasti: ARIMA (6, 0, 0) ja SARIMA (0, 0, 0)(1, 1, 0, 12). Julkaisussa käsiteltiin useita eri LSTM-malleja, mutta tuloksien

vertailukelpoisuutta varten niistä taulukoitiin normaali LSTM-malli. Datasta mainittiin kuukausittaisen myyntivolyymien olevan noin 500–1500 välillä, joten RMSE arvojen voidaan olettaa olevan suurempia.

Shetty et al. (2022) vertailivat ARIMA, SARIMA sekä LSTM ennustetarkkuuksia 2016–2019 välillä tehdyistä autokaupoista viikkotasolla. Datapisteitä oli lopulta 194, jotka jaettiin 158 pistettä koulutukseen, 20 validaatioon sekä 15 mallien testaukseen. Dataa ei muokattu stationääriseksi aluksi ja testit toteutettiin viikkotason tarkkuudella. Valitut metriikat olivat MAE sekä RMSE. ARIMA:n parametreiksi muodostuivat (2, 1, 3) ja SARIMA:n (2, 1, 2)(1, 1, 0, 52). LSTM-malli koostui kuudesta kerroksesta ja siinä käytettiin Nadam-mallia optimoimaan parametrejä.

Li et al. (2023) tutkivat ennustemallinnusta neljän kiinalaisen supermarketin myyntidatalla 1.1.2017–1.8.2021 väliltä. Data sisälsi yli 2 miljoonaa havaintoa sekä yli 3000 tuotetta, mutta mallinnukseen valikoitiin kurkun sekä banaanien myynti. Ennustetarkkuus toteutettiin päivätasolla 30 päivää eteenpäin. Valitut metriikat olivat MAPE sekä RMSE. Aineisto muokattiin ARIMA mallia varten stationääriseksi differensoinnilla mallin d parametrin avulla. Julkaisussa oli lisäksi malli, joka sisälsi exogeenisiä muuttujia kuten sään vaikutuksen myyntiin. Tarkempia ARIMA:n sekä LSTM:n parametrejä ei mainittu ollenkaan eikä ristivalidaation metodeista.

Boukrouh et al. (2025) tutki hybridimallin käyttöä myynnin ennustamisessa ja sen data oli kerätty 2.1.2011–30.12.2023 väliltä supermarketista USA:ssa. Data skaalattiin max-min normalisoinnilla ja 80 % säilytettiin koulutus- sekä 20 % testidataan. Valitut metriikat olivat ME, MAE, RMSE sekä MAPE. Mallit koulutettiin ja testattiin päivätason datalla. Julkaisussa ei ollut ARIMA mallia, mutta sen tilalla oli SARIMA (2, 1, 1)(2, 0, 2, 7). LSTM-malli koostui kolmesta piilokerroksesta, joissa jokaisessa oli 50 neuronua. Se koulutettiin Adam-mallilla 0,001 oppimisnopeudella. Julkaisussa ei käsitelty ristivalidaatiota eikä tarkemmin kuvattu datan stationäärisyyttä. Ehdotettu hybridimalli oli tarkin malleista.

5.2 Ennustetarkkuuden vertailu eri aineistoilla

Kerätty aineisto muodostaa lähes yksimielisen mielipiteen LSTM-mallin paremmuudesta (Taulukko 5). Ainoastaan kahdessa tutkimuksessa ARMA-mallit saivat paremman

ennustetarkkuuden. Läheskään kaikista aineistoista ei löytynyt kaikkia työhön valittuja metriikoita, joten tyhjät ruudut ovat tummennettuina taulukossa 5. Aineisto sisältää kokonaisuudessaan 20 havaintoa yhdeksästä eri julkaisusta. Kaikista julkaisuista oli myös havaittavissa, että parempi malli oli jokaiselta metriikalta aina parempi.

Taulukko 5: Kerätyt tulokset

Tutkimus	Datasetti	RMSE (ARIMA)	MAPE (ARIMA)	MAE (ARIMA)	RMSE (LSTM)	MAPE (LSTM)	MAE (LSTM)
1 Sirisha et al. (2022)	Sales data	8,6812	0,0616	6,4811	3,9173	0,0299	3,2580
Sirisha et al. (2022)	Sales data	7,7274	0,0562	6,0108	3,9173	0,0299	3,2580
2 Abbasimehr et al. (2020)	Sales data	3531,78			2267,11		
3 Ensafi et al. (2022)	Sales data	282,50	0,35		137,22	0,1839	
Ensafi et al. (2022)	Sales data	205,68	0,29		137,22	0,1839	
4 Temür et al. (2019)	Sales data	16,7450	0,1210	12,8010	21,7570	0,1500	17,9530
5 Athiyarath et al. (2020)	Short-term Sales data	55,00			65,00		
Athiyarath et al. (2020)	Midterm Sales data	136,00			185,00		
Athiyarath et al. (2020)	Long-term Sales data	185,00			243,00		
6 Wang et al. (2021)	Sales data 1	3507,88	0,06		113,45	0,0020	
Wang et al. (2021)	Sales data 2	9108,20	30,15		272,00	0,2830	
Wang et al. (2021)	Sales data 3	1061,47	0,05		293,01	0,0121	
Wang et al. (2021)	Sales data 4	5647,35	0,12		116,25	0,0084	
Wang et al. (2021)	Sales data 5	10273,37	0,08		139,87	0,0062	
7 Shetty et al. (2022)	Sales data	148,74		131,69	85,50		58,58
8 Boukrouh et al. (2025)	Sales data		0,4689			0,1412	
9 Li et al. (2023)	Sales data 1	123,69	6,0679		98,84	4,1142	
Li et al. (2023)	Sales data 2	55,27	0,4844		43,58	0,3591	
Li et al. (2023)	Sales data 3	35,82	0,3163		29,54	0,2741	
Li et al. (2023)	Sales data 4	80,59	0,4474		60,60	0,2538	

Kun vertaillaan havaintoaineiston RMSE-tuloksia, voidaan havaita niiden vaihtelevan alle 10 ja yli 10 000 välillä. Tämän havainto on selitettävissä kaavan (3) luonteen takia. Aineistot sisälsivät hyvin eriäviä suuruusluokkia havaintopisteiden arvoissa, mikä saa aikaan RMSE-luvun vaihtelua julkaisujen välillä. MAE-tulokset riippuvat yhtä lailla havaintopisteiden suuruudesta. Sitä ei kuitenkaan löytynyt, kun kolmesta julkaisusta valitulla aineistolla.

MAPE-tulokset ovat paremmin vertailukelpoisia, sillä se antaa virheelle prosentuaalisen arvon. Hyvä MAPE on alle 10 % ja sen saavutti lähes jokainen LSTM-malli. ARMA-mallit eivät saaneet alle 10 % MAPE arvoja yhtä usein, kun LSTM-mallit. Ainoa kerätty julkaisu, jossa ARIMA oli parempi ja siitä löytyi MAPE, ei saavuttanut alle 10 % tarkkuutta. MAPE-arvoista on kuitenkin havaittavissa sen herkkyyden outlier-arvojen kanssa. Tämä on havaittavissa Wang et al. (2021) julkaisun 3000 % (ARIMA) virheprosentissa.

Myynnin sekä kysynnän ennustamisessa MAE sekä MAPE ovat helpommin tulkittavia päätokeksenteon tukena. MAE tarjoaa konkreettisen poikkeaman valitussa yksikössä kuten kap-palemäärissä tai euroissa. MAPE puolestaan on helposti tulkittavissa eri skaalaisten datojen

välillä prosenttiyksikön avulla. Kumpaakin virhemittaria voidaan hyödyntää varastonhallinnan sekä tuotannon suunnittelussa käytännön tasolla. Esimerkiksi Sirisha et al. (2022) julkaisusta voitaisiin päätellä tuotannon tarvitsevan sisältää neljän tuotteen verran puskuria kuukausitasolla. Mikäli MAE sekä MAPE olisivat olleet suurempia, olisi varastointikustannusten arviointi sekä toimitusketjun suunnittelu haastavampaa.

RMSE korostaa virheitä enemmän verrattuna MAE. Näin ollen sitä voidaan soveltaa tuotteisiin, joiden kysyntämäärät voivat vaihdella kriittisemmin. Mittarina RMSE soveltuu parhaiten suotuisamman ennustemallin valitsemiseen herkkyytensä vuoksi. Pienempi RMSE on siis suoraan yhteydessä vähemmän poikkeamille alttiille ennusteelle.

5.3 Tuloksien vertailukelpoisuuden haasteet

Vaikka kerätyt julkaisut sisälsivät vähintään ARIMA- sekä LSTM-mallin, niiden vertailukelpoisuus muodostaa haasteita samoista mittareista huolimatta. Merkittävin havaittu haaste oli julkaisujen läpinäkyvyyden puutteessa. Kaikista julkaisuista ei löytynyt tarkempaa kuvausta LSTM-mallin hyperparametrien valikoinnista, mikä heikentää tulosten yleistettävyyttä sekä toistettavuutta. Julkaisujen välillä on hyödynnetty eri metriikoita tähän työhön valittujen RMSE, MAE sekä MAPE:n lisäksi. MAPE on näistä mittareista ainoa, jota voidaan hyödyntää eri julkaisujen välillä, sillä se antaa virheen prosentuaalisena arvona. MAPE ei ole mittarina kuitenkaan täydellinen erilaisten datasettien välillä. Tämä on havaittavissa, kun datasetit ovat jakautuneet eritavoin. Wang et al. (2021) julkaisun arvot ovat hyvin suuria ja vastaavasti Sirisha et al. (2022) hyvin pieniä. Näin ollen prosentuaalinen virhe on ennustetussa myynnin määrässä hyvin poikkeava kappalemäärissä.

Julkaisujen datasetit sisälsivät vähintään noin 200 havaintoa. Enimmillään havaintopisteitä oli yli 2 miljoonaa. LSTM-malli suosii suurempia aineistoja verrattuna ARIMA-malleihin, joka voi johtaa puolueellisuuteen LSTM-mallin eduksi. Tämä aiheutuu LSTM-mallin luonteesta, sillä se koostuu useista matemaattisista yksiköistä. Nämä yksiköt tarvitsevat koulutusta varten merkittävän määrän dataa, jotta LSTM-yksiköiden lopputulos saadaan mallintamaan koulutusvaiheessa syötettyä dataa. Julkaisut käsitelivät myös puutteellisesti LSTM-mallin tarkkuuden poikkeamia koulutusdatan, testi- sekä validaatiodatan välillä. Muodostettu malli voi olla tarkka, mutta samalla olla ylisovitettu. Siksi olisi tärkeätä tarkastella ennustetarkkuuksia testidatan lisäksi validaatiodatalla. Kaikissa julkaisuissa ei myöskään

käsitelty ristivalidaation metodeja tai hyödynnetty erityisesti aikasarjadataalle suunnattuja metodeja.

Datan esikäsitteilyllä on vaikutus mallien tuottamiin lopputuloksiin. Useassa julkaisussa käsiteltiin tilastollisista malleista ARIMA-mallia, joka pystyy käsittelemään trendiä differensoinnilla. ARIMA ei kuitenkaan kykene kausivaihtelun käsittelyyn. Kausivaihtelua sisältävät datasetit johtavat puolueellisiin tuloksiin LSTM-mallin eduksi, sillä se kykenee käsittelemään epästationääristä dataa. Osassa julkaisuja käytettiin SARIMA-mallia, joka pystyy kuitenkin käsittelemään kausivaihtelun sekä trendin. ARIMA:n osalta yksikään julkaisu ei käsitellyt tulosten jäännösarvojen tarkastelua, vaikka niille on oletuksena normaalijakautuneisuus. Jäännösarvojen autokorrelaatioiden käsittely puuttui myös lähes jokaisesta julkaisusta, vaikka valitut parametrit tuottivat julkaisuissa parhaat tulokset valittujen metriikoiden suhteen. Jäännösarvojen merkitys näkyy selkeämmin päivätason datalla, sillä sen muokkaaminen kuukausitasoille pehmittää vaihtelua. Täten vertailukelpoisuus heikkenee julkaisujen välillä, jotka ovat kuukausitasolla verrattuna viikko- tai päivittäiselle tasolle.

6 Johtopäätökset

Tässä kandidaatin työssä tutkittiin LSTM- sekä ARIMA-mallien hyödyntämistä kysynnän sekä myynnin ennustamisessa. Tutkimus toteutettiin systemaattisena kirjallisuuskatsauksena. Päättökysymys oli ”*1. Miten LSTM- vertautuu perinteisempiin ARMA-malleihin ennustemallinnuksessa?*”. Siihen tarkennuttiin alatutkimuskysymyksillä ”*1.1 Parantuuko ennustetarkkuus LSTM-mallilla?*” sekä ”*1.2 Miten mallien tulkittavuudet eroavat toisistaan?*”. Niiden avulla pyrittiin rakentamaan selkeä kuva mallien nykytilasta sekä suunnasta tulevaan.

1.1 Parantuuko ennustetarkkuus LSTM-mallilla?

Kerätyn aineiston pohjalta 16/20 testeistä tuotti LSTM-mallilla paremman ennustetarkkuuden kaikilla valituilla metriikoilla. Ennustetarkkuuksia mitattiin RMSE, MAE sekä MAPE metriikoilla. Tuloksiin vaikuttaviksi tekijöiksi havaittiin datan luonne, koko, ennustehorisontti sekä parametrien määrittely. Tulosten vertailukelpoisuudesta löydettiin haasteita samoista tekijöistä. Kaikista julkaisuista ei löytynyt selkeätä kuvausta datan stationäärisyydestä tai sen käsittelystä, valitusta ennustetarkkuudesta, datan kokoluokasta eikä mallien parametreista.

ARMA-mallit olivat tarkempia kahdessa julkaisussa. Temür et al. (2019) tekemässä julkaisussa aineiston koko oli ARMA-malleille suotuisampi ollessaan vain 124 havaintoa. Siinä MAPE:n arvo oli kuitenkin 12 %, joten mallia ei ole alle 10 % raja-arvon. Athiyarath et al. (2020) julkaisussa puolestaan ei mainittu havaintojen määrää, mutta siinä ARIMA oli paras malli lyhyellä, keskipitkällä sekä pitkällä ennustehorisontilla RMSE sekä selitysasteen mukaan. Selitysaste on ennustemalleissa huono mittari, sillä se kuvastaa selittävien muuttujien osuutta selitettävän muuttujan varianssista. Mikäli malli saa siis hyvin korkean selitysasteen, sen parametrit voivat olla ylisovitettuja. Julkaisun ei sisältänyt tarkempaa kuvausta parametreista sekä datan kokoluokasta, joten tulokset eivät ole välttämättä luotettavia.

Merkittävimmät erot ennustetarkkuuksissa oli havaittavissa Wang et al. (2021) julkaisussa, jossa LSTM-mallin MAPE oli alimmillaan 0,002 ja vastaavan datasetin ARIMA:n MAPE oli 0,06. Julkaisu sisälsi tuotteen, jonka kohdalla ARIMA:n MAPE sai 3015 % arvon ja LSTM sai puolestaan 28,3 % arvon. Julkaisu sisälsi yli 6000 havaintoa, joten LSTM-mallin

voidaan olettaa saavan parempia tuloksia. Yli 2800 % poikkeama voi kuitenkin viitata data-setin luonteeseen, sillä epästationäärinen data ei tuota luotettavia tuloksia ARIMA-mallilla. Wang et al. (2021) julkaisun pohjalta voidaan kuitenkin todeta LSTM-mallin tuottaneen paremmat tulokset kyseisten tuotteiden dataseiteillä.

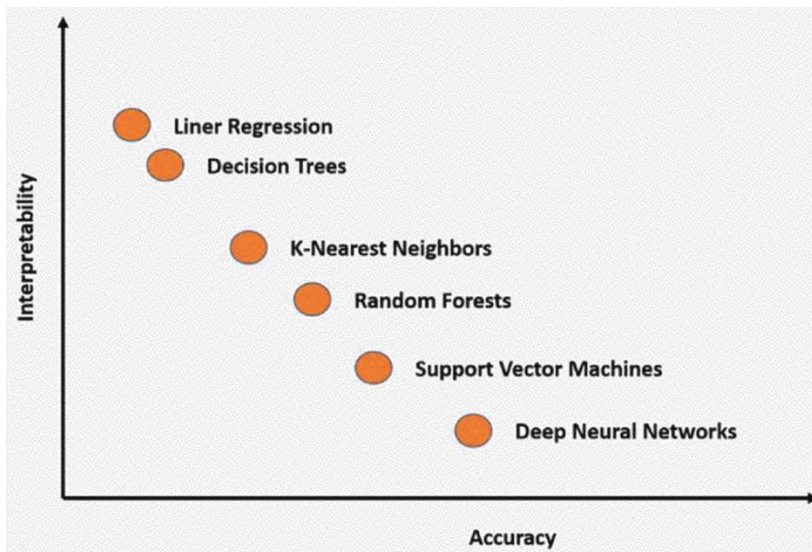
1.2 Miten mallien tulkittavuudet eroavat?

Kerättyjen julkaisujen pohjalta voidaan todeta ARIMA-mallin olevan selkeästi helpommin tulkittavissa. Lähes jokaisesta julkaisusta löytyi tarkat parametrit sekä selitykset ARIMA-mallien parametrien määrittelykselle. ARIMA-mallin tulkittavuus on konkreettisesti havaittavissa, kun sen matemaattinen malli kirjoitetaan auki. Valitut parametrit kuvaavat suoraan ennustettavaan arvoon vaikuttavien aikaisempien havaintojen määrää.

ARIMA-mallin tulkittavuuden haastekohdat ovat painojen määrittämisen ymmärtämisessä. Se vaatii tilastollisten menetelmien, kuten uskottavuusfunktioiden ymmärtämistä. Muutoin malli asettamien painojen ymmärtäminen jää vajavaiseksi. ARIMA:n kokonaisvaltainen ymmärtäminen mahdollistaa ennusteiden lisäksi niiden hyödyntämisen käytännötasolla. Mikäli mallia hyödynnetään ilman sitä, ei tulosten varmuudesta voi olla muuta kuin skeptinen.

LSTM-mallin hyperparametrien määrittelystä oli enemmän havaittavissa puutteita ja julkaisujen välillä oli havaittavissa eroavaisuuksia niiden valinnoista. Niillä voidaan määrittellä solutilojen sekä yksiköiden määrän lisäksi useita eri asioita. Mallin rakentuessa useista eri yksiköistä epävarmuus mallin tulkittavuudessa kasvaa, sillä havaintojen vaikutukset ennusteisiin eivät ole suoraan selitettävissä muutaman matemaattisen kaavan takaa. LSTM-mallin kouluttamisessa se voi käydä aineiston läpi halutun määrän Epoch-hyperparametrin määrittämisellä.

Tämä tulkittavuuden puute kulkee käsi kädessä kuvassa 8 havaittavasta suhteesta ennustetarkkuuteen. LSTM-malli olisi parhaiten tulkittavissa, jos sen hyperparametrien määrittäminen kuvattaisiin niiden arvojen lisäksi. Sen avulla voitaisiin toistaa julkaisuissa tehtyjä havaintoja ja vertailla eri aineistojen välillä mallien yleistettävyyttä.



Kuva 8: Mallien tarkkuuden ja tulkittavuuden suhde. (Nandi & Pal 2022)

ARIMA-mallien hyödyntämisessä tulkittavuuteen tuo lisähaasteita monimutkaisemmat mallit kuten SARIMA tai SARIMAX. Lisätyt parametrit lisäävät uusia osia yhtälöön sekä painoja määritettäväksi. ARIMA:n tulkittavuus heikkenee myös, mikäli malliin lisätään korkeasteisia parametreja, sillä yksittäisen AR tai MA parametrin vaikutuksen havainnointi voi olla vaikeampaa. LSTM-mallin painojen sekä bias-termien määrittäminen koulutusdatan pohjalta on kuitenkin aina haastavampaa verrattuna ARIMA-malliin tai sen johdannaisiin. Mallin tulkittavuus ei ole kuitenkaan itseisarvo hyvälle ennusteelle. Mallin käyttöön vaikuttaa kuitenkin kuka ennustaa ja mihin käyttötarkoitukseen. Tulkittavuuden sekä ennustetarkkuuden suhde on siis käyttäjän itse arvioitava, sillä liian monimutkaisesti tulkittavan ennusteen tarjoaminen väärälle kuulijalle ei tuo ymmärrettäviä tuloksia.

1. Miten LSTM- vertautuu perinteisempiin ARMA-malleihin ennustemallinnuksessa?

ARMA-mallien merkittävin hyöty on niiden yksinkertaisuudessa. Ne perustuvat autoregressioon sekä liukuvan keskiarvon hyödyntämiseen aikasarjadataan ennustemallinnuksessa. Mallit toimivat suhteellisen pienellä datalla, mutta vaativat sen olevan stationäärinen. Se on kuitenkin kierrettävissä ARMA-mallien vaihtoehtoisilla malleilla, kuten ARIMA sekä SARIMA. ARMA-malleista yleisin julkaisuissa käytetty malli oli ARIMA, joka mahdollistaa trendin poistamisen differensointi parametrilla. Mallit eivät kykene epälineaaristen riippuvuuksien käsittelyyn, joka voi heikentää ennustetarkkuuksia.

LSTM-malli pystyy käsittelemään epälineaarista dataa sekä mallintamaan pitkän aikavälin riippuvuuksia. Näin ollen se toimii myös paremmin epästationääriselle datalle. LSTM-

mallin kyvykkyys pohjautuu sen neuroverkkorakenteeseen, jossa aiemmat havainnot säilyvät pitkä sekä lyhytaikaisessa muistissa. Se toimii myös paremmin suuremmilla dataseiteillä verrattuna ARMA-malleihin. LSTM-mallit sisältävät enemmän parametreja verrattuna ARMA-malleihin. LSTM-mallin aktivaatiofunktioiden luonteen takia data voisi olla hyvä normalisoida. Se voi kuitenkin johtaa tärkeiden havaintojen katoamiseen datasta ja johtaa heikentyneeseen ennustetarkkuuteen.

ARMA-mallit vaativat vähemmän laskentatehoa verrattuna LSTM-malleihin. Tämä aiheutuu vähäisempien parametrien määrittelystä. ARMA-mallien parametrit ovat matemaattisesti johdettavissa. LSTM-mallit puolestaan käyttävät gradienttilaskentaa sen painokertoimien ja bias-termien määrittelyyn, joka vaatii merkittävästi enemmän laskennallista tehoa. LSTM-mallit koostuvat myös usein useasta eri yksiköstä, mikä lisää laskutapahtumien määrää.

Syväoppimismallit tuntuvat kehittyvän nopeammin verrattuna niistä kirjoitettuihin tieteellisiin julkaisuihin. Muutama vuotta sitten tehdyt julkaisut ovat jo melkein vanhaa tietoa. Jatkotutkimusmahdollisuuksia löytyy tämän työn rajauksen ulkopuolelta muun muassa hybridimallien kohdalla. Aineistossa useat julkaisut tarjosivat hybridimalleja, jotka sisälsivät koneoppimismalleja, syväoppimismalleja sekä tilastollisia malleja. Hybridimallit voivat tuottaa tarkempia tuloksia verrattuna yksittäisiin malleihin, mikä oli havaittavissa kerätyn aineiston perusteella. Kuitenkin usean eri mallin yhdistäminen voi johtaa kasvaneeseen tulkittavuuden puutteeseen, sillä ne vaativat usean eri mallin toiminnan yksinään sekä yhdessä. ARIMA:sta sekä LSTM-mallista koostuva hybridimalli voisi käytännön tasolla toimia siten, että ARIMA ennustaisi aikasarjaa yleisesti ja LSTM ennustaisi jäännösarvot. Näin hybridimallilla voitaisiin parantaa ennustetarkkuutta, kun jäännösarvojen kohdalla voi olla tilastollisilla malleilla haasteita. Usean eri mallin käyttäminen samanaikaisesti luultavasti johtaa myös laskentatehon kasvaneisiin vaatimuksiin sekä sen kuluttamaan energiaan.

Jatkotutkimusta voitaisiin myös toteuttaa lisäämällä malleihin eksogeenisiä muuttujia kuten säätiedot ja mainoskampanjat. Vaihtoehtoisesti malliin voitaisiin lisätä makrotalouden muuttujia kuten bruttokansantuotteen kasvuprosentti tai inflaatio. Ne voisivat lisätä ennustetarkkuutta sekä tulkittavuutta. Eksogeeniset muuttujat tulisi kuitenkin valita ennustettavalle muuttujalle sopivaksi. Niiden vaikutusta voidaan tarkastella esimerkiksi SARIMAX-mallin eksogeenisten muuttujien kertoimilla sekä niiden tilastollisella merkittävyydellä. Aineistojen kokoluokasta voitaisiin tehdä myös jatkotutkimusta, sillä LSTM-mallien kohdalla ei

löytynyt selkeitä vertaisarvioituja julkaisuja havaintojen määrän vaikutuksesta ennustetarkkuuteen sekä tulkittavuuteen.

Tämän työn rajoitteena oli mallien tarkasteleminen myynnin sekä kysynnän näkökulmasta. Vertailevia julkaisuja ei ollut saatavilla merkittävää määrää, joten aiheesta voidaan silti tehdä merkittävää tutkimusta jatkossa. Ennustamisen kohteita on myynnin sekä kysynnän ulkopuolella kuten kappaleessa 1.1 sekä 2.1 mainittiin. Syväoppimismallien yleistyessä jatkotutkimusta voitaisiin toteuttaa raportoinnin standardisoinnissa, sillä kerätyn aineiston pohjalta oli havaittavissa paljon eroavaisuuksia hyperparametrien valinnassa sekä ristivalidaatiossa. Tämä lisäisi mallien tulkittavuutta, sillä julkaisujen tuloksia saataisiin toistettua paremmin.

Lähteet

- Abbasimehr, H., Shabani, M. & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering*. Vol. 143, Artikkelin numero 106435.
- Athiyarath, S., Paul, M. & Krishnaswamy, S. (2020). A Comparative Study and Analysis of Time Series Forecasting Techniques. *SN Computer Science*. Vol. 1(3), Artikkelin numero 175.
- Auffarth, B. (2021). *Machine Learning for Time-Series with Python : Forecast, Predict, and Detect Anomalies with State-of-the-Art Machine Learning Methods*. 371 s. Birmingham: Packt Publishing, Limited.
- Bergmeir, C., Hyndman, R. J. & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*. Vol. 120, s. 70–83.
- Bianco, A. M., García Ben, M., Martínez, E. J. & Yohai, V. J. (2001). Outlier Detection in Regression Models with ARIMA Errors using Robust Estimates. *Journal of Forecasting*. Vol. 20(8), s. 565–579.
- Boukrouh, I., Idiri, S., Tayalati, F., Azmani, A. & Bouhsaien, L. (2025). A Hybrid Approach for Sales Forecasting: Combining Deep Learning and Time Series Analysis. *International Journal of Engineering*. Vol. 38(4), s. 859–870.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time series analysis forecasting and control* (Rev. ed.). 575 s. San Francisco: Holden-Day.
- Box, G. E. P. & Tiao, G. C. (1975). Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*. Vol. 70(349), s. 70–79.
- Carvalho, A., Lefèvre, S., Schildbach, G., Kong, J. & Borrelli, F. (2015). Automated driving: The role of forecasts and uncertainty—A control perspective. *European Journal of Control*. Vol. 24, s. 14–32.

- Cerqueira, V., Torgo, L. & Mozetič, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*. Vol. 109(11), s. 1997–2028.
- Cipra, T. (2020). *Time Series in Economics and Finance*. 410 s. Cham: Springer International Publishing.
- Cissoko, M. B. H., Castelain, V. & Lachiche, N. (2025). Multi-Way adaptive Time Aware LSTM for irregularly collected sequential ICU data. *Expert Systems with Applications*. Vol. 261, Artikkel 125548.
- Dally, W. J., Keckler, S. W. & Kirk, D. B. (2021). Evolution of the Graphics Processing Unit (GPU). *IEEE Micro*. Vol. 41(6), s. 42–51.
- Ensafi, Y., Amin, S. H., Zhang, G. & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*. Vol. 2(1), Artikkel 100058.
- Euroopan komissio. (2024). The cost of uncertainty – new estimates. [Verkköjulkaisu]. [Viitattu 1.2.2025]. Saatavilla: https://economy-finance.ec.europa.eu/economic-forecast-and-surveys/economic-forecasts/autumn-2024-economic-forecast-gradual-rebound-adverse-environment/cost-uncertainty-new-estimates_en
- Fildes, R., Ma, S. & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*. Vol. 38(4), s. 1283–1318.
- Gao, H., Kou, G., Liang, H., Zhang, H., Chao, X., Li, C. & Dong, Y. (2024). Machine learning in business and finance: a literature review and research opportunities. *Financial Innovation*. Vol. 10(1), Artikkel 86.
- Gautam, A. & Singh, V. (2020). Parametric versus non-parametric time series forecasting methods: A review. *Journal of Engineering Science and Technology Review*. Vol. 13(3), s. 165–171.
- Gilliland, M., Tashman, L., Sglavo, U., Makridakis, S. G. & Petropoulos, F. (2021). *Business Forecasting: The Emerging Role of Artificial Intelligence and Machine Learning*. 414 s. Newark: John Wiley & Sons, Incorporated.

- Goh, C. (2024). Analysts' Earnings per Share Forecasts: The Effects of Forecast Uncertainty and Forecast Precision on Investor Judgements. *Abacus (Sydney)*. Vol. 60(1), s. 172–204.
- Hewamalage, H., Ackermann, K. & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*. Vol. 37(2), s. 788–832.
- Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*. Vol. 9(8), s. 1735–1780.
- Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. Vol. 06(02), s. 107–116.
- Jin, Y., Cao, Q., Wang, K., Zhou, Y., Cao, Y. & Wang, X. (2023). Prediction of COVID-19 Data Using Improved ARIMA-LSTM Hybrid Forecast Models. *IEEE Access*. Vol. 11, s. 67956–67967.
- Jo, T. (2021). *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. 391 s. Cham: Springer International Publishing.
- Khadse, V. M., Mahalle, P. N. & Shinde, G. R. (2020). Statistical Study of Machine Learning Algorithms Using Parametric and Non-Parametric Tests: A Comparative Analysis and Recommendations. *International Journal of Ambient Computing and Intelligence*. Vol. 11(3), s. 80–105.
- Khalkar, R., Dikhit, A. & Goel, A. (2021). Handwritten Text Recognition using Deep Learning (CNN & RNN). *IARJSET*. Vol. 8, s. 870–881.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and modelselection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. Vol. 2. Montreal, Kanada: Morgan Kaufmann Publishers Inc. s. 1337–1143
- Koutsandreas, D., Spiliotis, E., Petropoulos, F. & Assimakopoulos, V. (2022). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*. Vol. 73(5), s. 937–954.
- Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. *Sociological Methods & Research*. Vol. 33(2), s. 188–229.

- Laome, L., Adhi Wibawa, G. N., Raya, R., Makkulau & Asbahuna, A. R. (2021). Forecasting time series data containing outliers with the ARIMA additive outlier method. *Journal of Physics.Conference Series*. Vol. 1899(1), Artikkel 12106.
- Leffrang, D. & Müller, O. (2024). Visualizing Uncertainty in Time Series Forecasts: The Impact of Uncertainty Visualization on Users' Confidence, Algorithmic Advice Utilization, and Forecasting Performance. *Journal of Forecasting*.
- Li, Z., Yu, W., Zhu, W., Wan, H., Peng, J. & Wang, H. (2023). Research on Fresh Produce Sales Prediction Algorithm for Store Based on Multidimensional Time Series Data Analysis. *EAI Endorsed Transactions on Scalable Information Systems*. Vol. 11(2), s. 1–11.
- Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel, Switzerland)*. Vol. 23(1), Artikkel 18.
- Liu, S. & Zhou, D. J. (2024). Using cross-validation methods to select time series models: Promises and pitfalls. *British Journal of Mathematical & Statistical Psychology*. Vol. 77(2), s. 337–355.
- Lütkepohl, H. & Kräzig, M. (2004). *Applied Time Series Econometrics*. 323 s. Cambridge: Cambridge University Press.
- Makridakis, S., Hibon, M. & Moser, C. (1979). Accuracy of Forecasting: An Empirical Investigation. *Journal of the Royal Statistical Society, Series a (General)*. Vol. 142(2), s. 97–145.
- Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS One*. Vol. 13(3), Artikkel e0194889.
- Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*. Vol. 36(1), s. 54–74.
- Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*. Vol. 38(4), s. 1346–1364.

Masters in Data Science (n.d.). What is ARIMA modeling? [Verkköjulkaisu]. [Viitattu: 25.3.2025]. Saatavilla: https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arma-modeling/?utm_source=chatgpt.com

Mbonyinshuti, F., Nkurunziza, J., Niyobuhungiro, J. & Kayitare, E. (2024). Health supply chain forecasting: a comparison of ARIMA and LSTM time series models for demand prediction of medicines. *Acta Logistica*. Vol. 11(2), s. 269–280.

Mehtab, S., Sen, J. & Dutta, A. (2021). Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models. Teoksessa: Thampi, S. M., Piramuthu, S., Li, K. C., Berretti, S., Wozniak, M. & Singh, D. (toim). *Machine Learning and Metaheuristics Algorithms, and Applications*. Singapore: Springer, s. 88–106.

Mienye, I. D., Swart, T. G. & Obaido, G. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information (Basel)*. Vol. 15(9), Artikkel 517.

Mignon, V. (2024). An Introduction to Time Series Models. Teoksessa: Mignon, V. (toim.). *Principles of Econometrics*. Switzerland: Springer. s. 287–349.

Nandi, A. & Pal, A. K. (2022). Interpreting machine learning models: learn model interpretability and explainability methods. 355 s. New York, New York: Apress.

Niako, N., Melgarejo, J. D., Maestre, G. E. & Vatcheva, K. P. (2024). Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with ARIMA and LSTM. *BMC Medical Research Methodology*. Vol. 24(1), s. 320–32.

Oliveira, P., Fernandes, B., Analide, C. & Novais, P. (2021). Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities. *Electronics (Basel)*. Vol. 10(10), s. 1–22.

Piccialli, F., Giampaolo, F., Prezioso, E., Camacho, D. & Acampora, G. (2021). Artificial intelligence and healthcare: Forecasting of medical bookings through multi-source time-series fusion. *Information Fusion*. Vol. 74. s.1–16.

Scopus. (2025). Elsevier. [Viitetietokanta]. [Viitattu: 3.4.2025]. Saatavilla: <https://www.scopus.com>

- Seaman, B. (2018). Considerations of a retail forecasting practitioner. *International Journal of Forecasting*. Vol. 34(4), s. 822–829.
- Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica. D*. Vol. 404, Artikkel 132306.
- Shetty, S. K. & Buktar, R. (2022). A comparative study of automobile sales forecasting with ARIMA, SARIMA and deep learning LSTM model. *International Journal of Advanced Operations Management*. Vol. 14(4), s. 366–387.
- Singh, U., Tamrakar, S., Saurabh, K., Vyas, R. & Vyas, O. P. (2023). Hyperparameter Tuning for LSTM and ARIMA Time Series Model: A Comparative Study. Teoksessa: Singh, U., Tamrakar, S., Saurabh, K., Vyas, R. & Vyas, O. P. *2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON)*. Mysore, Intia: IEEE. s. 1–6.
- Sirisha, U. M., Belavagi, M. & Attigeri, G. (2022). Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison. *IEEE access*. Vol. 10, s. 124715–124727.
- Statista. (2024). Market size of machine learning (ML) market globally from 2020 to 2030. [Verkköjulkaisu]. [Viitattu: 31.1.2025]. Saatavilla: <https://www.statista.com/forecasts/1449854/machine-learning-market-size-worldwide>
- Temür, A. S., Akgün, M. & Temür, G. (2019). Predicting Housing Sales in Turkey Using ARIMA, LSTM and Hybrid Models. *Journal of Business Economics and Management*. Vol. 20(5), s. 920–938.
- Tsay, R. S. (2000). Time Series and Forecasting: Brief History and Future Research. *Journal of the American Statistical Association*. Vol. 95 (450), s. 638–643.
- Vamsikrishna, A. & Gijo, E. V. (2024). New Techniques to Perform Cross-Validation for Time Series Models. *Operations Research Forum*. Vol. 5(2), Artikkel 51.
- Wang, C., Chien, C. & Trappey, A. J. C. (2021). On the Application of ARIMA and LSTM to Predict Order Demand Based on Short Lead Time and On-Time Delivery Requirements. *Processes*. Vol. 9(7), Artikkel 1157.

World Economic Forum. (2025). The Future of Jobs Report 2025. [Verkköjulkaisu]. [Viitattu: 31.1.2025]. Saatavilla: <https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/>

Yadav, A. (2024). A Comparative Study of Time Series, Machine Learning, and Deep Learning Models for Forecasting Global Price of Wheat. *Operations Research Forum*. Vol. 5(4), Artikkele 113.

Yan, Z., Appelbaum, D., Kogan, A. & Vasarhelyi, M. A. (2023). Teaching Predictive Audit Data Analytic Techniques: Time-Series Forecasting with Transactional and Exogenous Data. *Journal of Emerging Technologies in Accounting*. Vol. 20(1), s. 169–194.