

Juho Naatula

# LSTM- JA TRANSFORMER-MALLIEN VERTAILU OSAKEKURSSIEN ENNUS- TAMISESSA

Kandidaatintutkielma  
Informaatioteknologian ja viestinnän tiedekunta  
Tarkastaja: Tapio Elomaa  
Kesäkuu 2024

# TIIVISTELMÄ

Juho Naatula: LSTM- ja Transformer-mallien vertailu osakekurssien ennustamisessa  
Kandidaattitutkielma  
Tampereen yliopisto  
Tietotekniikan tutkinto-ohjelma  
Kesäkuu 2024

---

Osakemarkkinat ovat yksi nykyaikaisen talouden kulmakivistä ja niiden käyttäytymisen oikein ennustaminen on mahdollisesti hyvin tuottoisaa. Lukuisat yritykset, sijoitusrahastot sekä yksilöt yrittävät jatkuvasti löytää uusia mahdollisia tapoja olla askeleen edellä markkinoita, jotta he voisivat minimoida sijoitusriskejä ja maksimoida mahdolliset tuotot. Osakemarkkinoiden ennustaminen on kuitenkin hyvin monimutkaista, ja siihen kehitetään ja sovelletaan jatkuvasti uusia teknologioita ja teknologioita. Uusimpia tähän käytettyjä teknologioita ovat erilaiset syväoppimismallit, kuten LSTM- ja Transformer-pohjaiset arkkitehtuurit. Tutkimuksessa verrataan LSTM-malleja Transformer-malleihin osakemarkkinoiden ennustamisessa.

Tämä tutkimus on toteutettu kirjallisuuskatsauksena ja jakautuu kolmeen pääosaan. Ensimmäisen osan tavoitteena on selittää miten syväoppimismallit, kuten LSTM-mallit ja Transformer-mallit, toimivat yleisellä tasolla. Toisessa osassa selitetään yleisimmät käytännöt osakemarkkinoiden ennustuksessa syväoppimismalleilla, kuten ennustuksen kohteet ja mallien tehokkuuden vertailu. Lopuksi vertaillaan eri tutkimuksissa käytettyjä malleja, niiden käyttötarkoituksia sekä tutkimuksien tuloksia.

Tutkimusten tuloksista käy ilmi, että Transformer-mallit ovat yleisesti parempia käsittelemään aikasarjadataa kuin perinteisemmät LSTM-mallit, vaikka LSTM-mallit on kehitetty juuri tähän tarkoitukseen, mutta tulos ei ole yksiselitteinen. Yleisesti tutkimusten tuloksissa Transformer-mallien vuosittainen tuotto on suurempi sekä tasaisempi kuin LSTM-malleilla, mutta tähän löytyy poikkeuksia. Monissa tutkimuksissa Transformer-mallit ovat myös hyvin tärkeitä poimimaan osakekohtaista optimismia tekstipohjaisesta datasta, kuten uutisista tai kommenteista. Mallin valitseminen on kuitenkin vain osa ennustamisprosessia ja monet muut tekijät, kuten laadukkaan datan käyttäminen ja oikean ennustuksen kohteen valitseminen, ovat myös hyvin tärkeä osa ennustamisprosessia.

Avainsanat: LSTM, Transformer, osakemarkkinoiden ennustaminen, tekoäly, koneoppiminen

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

# ABSTRACT

Juho Naatula:  
Comparison of LSTM Models and Transformer Models in Stock Price Forecasting.  
Bachelor's thesis  
Tampere University  
Information Technology  
June 2024

---

Stock markets are one of the cornerstones of modern economy, and correctly predicting their behavior can be potentially very lucrative. Numerous companies, investment funds, and individuals are constantly trying to find new possible ways to stay ahead of the market, so that they can minimize investment risk and maximize potential returns. Predicting stock markets is very complex and new techniques and technologies are continuously being developed and applied. The latest technologies used for this are various deep learning models. This research compares LSTM models with Transformer models in stock market prediction.

This study is conducted as a literature review and is divided into three main sections. The first part aims to explain, at a general level, how deep learning models, like LSTM and Transformer models, work. The second part outlines common practices in stock market prediction using deep learning models, such as prediction targets and model efficiency comparisons. Lastly, models used in various studies, their purposes, and the research results are compared.

The results of the studies reveal that Transformer models are generally better at handling time-series data than the more traditional LSTM models, even though LSTM models were specifically developed for this purpose, but the result isn't unambiguous. Usually, the annual returns of Transformer models are higher and more stable than those of LSTM models, but there are exceptions to be found. In many studies, Transformer models are also very important for capturing stock-specific optimism from text-based data like news or comments. However, choosing a model is just one part of the forecasting process, and many other factors, such as high-quality data and selecting the right prediction target, are also very important parts of the process.

Keywords: LSTM, Transformer, stock market prediction, artificial intelligence, machine learning

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

# SISÄLLYSLUETTELO

1. JOHDANTO .....	1
2. NEUROVERKOT .....	3
2.1 Syvät neuroverkot.....	3
2.2 LSTM-mallit ja muisti .....	4
2.3 Transformer-mallit ja huomio .....	5
2.4 LSTM-verkot verrattuna Transformer-verkkoihin .....	7
3. OSAKEMARKKINOIDEN ENNUSTAMINEN .....	8
3.1 Osakemarkkinoiden data .....	8
3.2 Neuroverkkojen käyttökohteet .....	8
3.3 Ennustustarkkuuden mittaaminen.....	8
4. MALLIEN VERTAILUTULOKSET .....	10
4.1 Käytettyjen mallien esittely.....	10
4.2 Mallien tulosten vertailu .....	14
YHTEENVETO.....	23
LÄHTEET .....	24

# LYHENTEET JA MERKINNÄT

AR	Annual Return, vuosittainen tuotto prosentti
GPT	Generative Pretrained Transformer
LSTM	Long Short Term Memory
MAPE	Mean Absolute Percentage Error, keskimääräinen absoluuttinen prosenttinvirhe
MDD	Maximum Draw Down
MSE	Mean Squared Error, keskineliövirhe
RNN	Recurrent Neural Network

# 1. JOHDANTO

Osakemarkkinoita on yritetty ennustaa niin kauan kuin ne ovat olleet olemassa. Ennustamisessa käytetty teknologia on kuitenkin kehittynyt vuosikymmenten varrella. Eräs ennustuksessa käytetty teknologia on tekoäly- ja syväoppimismallit, joiden käyttö on yleistynyt huomattavasti 2010-luvulla (Wang & muut 2022, s.1). Syväoppimismallit ovat erityisen sopivia osakemarkkinoiden ennustamisessa, sillä osakemarkkinat ovat erittäin monimutkaisia ja tekoälymallit pystyvät hallitsemaan jopa vaikeat datayhteydet (Wang & muut 2022, s.1). Lisäksi osakemarkkinoilla on hyvin paljon dataa, jota voidaan käyttää syväoppimismallien opettamisessa (Li & muut 2023).

Osakemarkkinoiden data on luonnostaan aikasarjadataa, joten aikasarjadataan kehitelty syväoppimismallit, kuten Hochreiter ja Schmidhuberin (1997) kehittelemät LSTM-mallit, ovat olleet luonnollinen valinta jo pitkän aikaan. Viime aikoina uusi sekvenssidataan keskittynyt Transformer-malli on saanut paljon huomiota ja se suoriutuu hyvin tehtävissä, kuten tekstin käsittelyssä, joihin on aiemmin käytetty RNN- ja LSTM-malleja (Gomez & muut 2017).

Tässä tutkimuksessa selitetään miten LSTM- ja Transformer-mallit toimivat yleisellä tasolla, mitä osakemarkkinoiden ennustamiseen liittyy ja miten ennustamistarkkuutta mitataan. Lopuksi yritetään selvittää, miten malleja käytetään osakemarkkinoilla ja miten niiden ennustamistarkkuus vertaasi toisiinsa.

Tutkimus on toteutettu kirjallisuuskatsauksena, jossa lähteet on haettu Andor sivuilta hakulauseella:

```
(financ* OR stock* OR crypto* OR currency* OR market*) AND ((Transformer* OR *Transformer) OR ((Transformer* OR *Transformer) AND attention*)) NOT (solar OR electric*).
```

Haut rajattiin vertaisarvioituihin tutkimuksiin, joissa julkaisuvuosi oli 2022–2024, jotta saataisiin uusia tuloksia. Multimedia Systems niminen lehti rajattiin pois, sillä kyseinen lehti ja sen tutkimukset eivät vaikuttaneet sopivan haettuun aiheeseen. Löydetyt tutkimukset rajattiin pidemmälle kriteerillä, että niissä on suora vertailu LSTM- ja Transformer-mallien välillä ja molemmat mallit käyttivät samaa dataa osakekurssien ennustuksessa. Esimerkiksi tutkimukset, joissa LSTM-malli käytti osakkeen hintadataa, kun taas Transformer-malli käytti uutisdataa, hylättiin. Tutkimuksissa vertailuasteikkona täytyi olla joko mallien tuottavuus tai ennustusvirheen MSE tai MAPE. Myös tutkimukset, joissa ennustettiin esimerkiksi säätä, jätettiin pois. Karsinnan jälkeen jäi seitsemän tutkimusta

mallien vertaamiseen ja muut lähteet on valittu tekoälymallien tai osakemarkkinoiden selventämiseen.

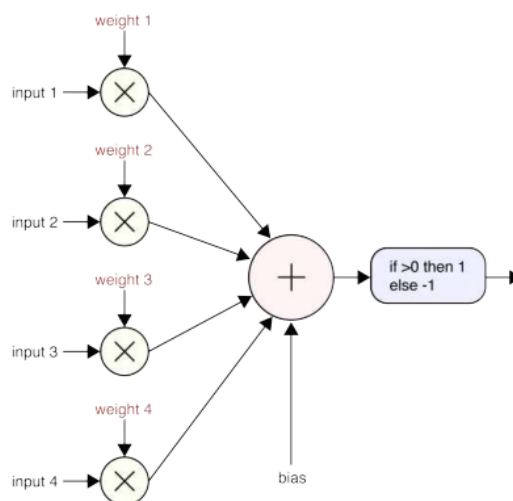
Luvussa 2. käydään läpi tarkemmin tekoäly algoritmien toimintaperiaatteita ja miten LSTM- ja Transformer-mallit toimivat. Sen jälkeen siirrytään tutkimaan tarkemmin, mihin LSTM- ja Transformer-mallit ovat erikoistuneet ja miten ne eroavat toisistaan. Luvussa 3 katsotaan tarkemmin osakekurssien ennustamista, millaista dataa ennustamiseen käytetään, miten ennustamista mitataan ja miten mallit sopivat osakekurssien ennustamiseen. Luvussa 4 tehdään kirjallisuuskatsaus valittuihin papereihin ja verrataan tuloksia, jotta saadaan selville, miten LSTM- ja Transformer-mallit vertaavat toisiinsa osakekurssien ennustamisessa. Lopuksi katsotaan yhteenveto tuloksista ja keskustellaan niistä vedettävistä johtopäätöksistä.

## 2. NEUROVERKOT

*Neuroverkot* ovat matemaattisia malleja, jotka oppivat muuntamalla mallin sisäisiä parametrejä, kuten neuroneiden painoja ja vinouma-arvoja opetusdatan mukaan. Ohjatussa oppimisessa pyritään opettamaan mallia tuottamaan sisääntuloarvoista ulostuloarvo, joka vastaa tiedossa olevaa oikeaa ulostuloarvoa. Saadun ulostuloarvon ja halutun ulostuloarvon välistä eroa kutsutaan virheeksi. Mallin oppimisen myötä virhe pienenee, eli mallin tuottama ulostuloarvo lähenee haluttua ulostuloarvoa. Malli opetetaan opetusdatalla syöttämällä sille sisääntuloarvot ja vastaavat halutut ulostuloarvot. Mallin opettamisen jälkeen mallia testataan testidatalla, josta malliin syötetään ainoastaan sisääntuloarvot ja mallin tuottamat ulostuloarvot verrataan haluttuihin ulostuloarvoihin ja niiden välinen virhe mitataan. Jos malli oppii opetusdatan ja toimii hyväksyttävästi testidatan kanssa, niin voidaan olettaa, että malli on oppinut jonkinlaisen tilastollisen kytköksen opetusdatasta ja pystyy ennustamaan halutun ulostulon datasta, jota malli ei ole nähnyt. (Szeliski 2022, s. 239)

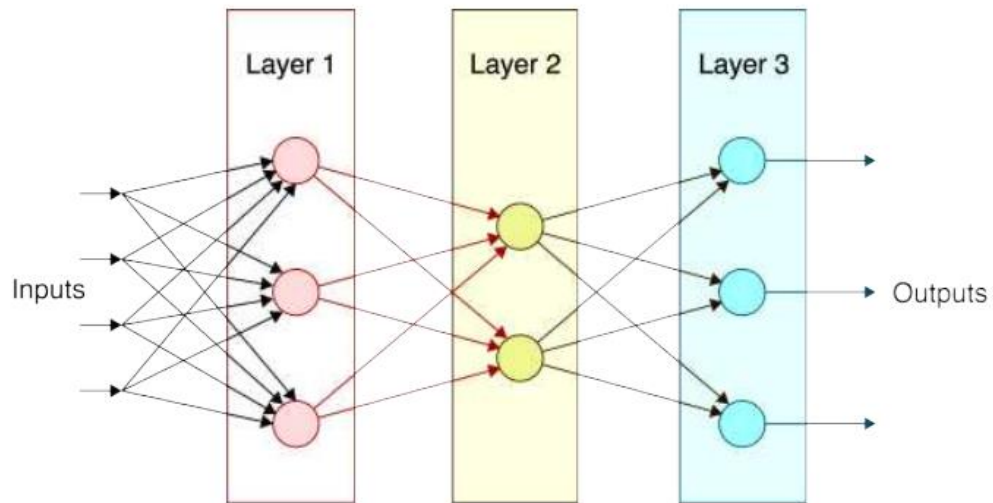
### 2.1 Syvät neuroverkot

Syvät neuroverkot koostuvat monista kerroksista neuroneita ja täysin kytketyissä neuroverkoissa jokainen neuroni vastaanottaa painotetun arvon jokaisesta edellisen kerroksen neuronista. Neuronin arvo on yhteenlaskettu summa kaikista painotetuista edellisen kerroksen arvoista, johon lisätään neuronin vinouma-arvo. Summattu arvo muunnetaan epälineaarisen aktivaatiofunktion läpi, joka takaa, että neuroverkko pystyy muotoilemaan epälineaarisia suhteita. (Szeliski 2022, s. 270–271)



**Kuva 1.** yksittäisen neuronin toimintaperiaate (Szeliski 2022, s. 270)

Kuva 1 kuvastaa yksittäistä neuronia, joka vastaanottaa 4 arvoa kuvastaen myös arvoihin kohdistuvia operaatioita. Jokainen neuroniin sisääntuloarvo painotetaan niiden omilla painoarvoilla, arvot summataan yhteen ja lopputulokseen lisätään vinouma-arvo.



**Kuva 2. täysin kytketty neuroverkko (Szeliski 2022 s. 271)**

Kuva 2 kuvastaa tavallisen täysin kytketyn neuroverkon toimintaa, jossa kaikki sisään tulevat arvot menevät koko neuroverkon läpi, ja viimeinen neuronikerros tuottaa ulostuloarvot. Tavalliset syvät neuroverkot ovat sopiva työkalu dataan, jossa datapisteet eivät ole yhteydessä toisiinsa. Kuitenkaan ne eivät ole sopivia aikasarjadataan, jossa monet datapisteet ovat yhteydessä toisiinsa ja joiden välisestä suhteesta voidaan kerätä ylimääräistä tietoa. Aikasarjadatan ennustamiseen on kehitelty omanlaisia neuroverkko-arkkitehtuureja.

## 2.2 LSTM-mallit ja muisti

*RNN (Recurrent Neural Network)* verkot ovat erikoistuneet dataan, jossa edelliset datapisteet vaikuttavat seuraaviin datapisteisiin. Yleisin tapaus on aikasarjadata, jossa edellisten päivän data on yhteydessä nykyisen päivän dataan ja näiden datapisteiden välinen suhde antaa meille enemmän tietoa, kuin molempien datapisteet erikseen. Sen sijaan, että *RNN*-verkoille syötetään vain yksi datapiste kerrallaan, ne käsittelevät kokoelman datapisteitä kerrallaan. *RNN*-verkot käsittelevät datapistekokoelman osissa ja jokainen osa jakaa dataa seuraaviin osiin. (Szeliski 2022, s. 321)

*RNN*-verkot toimivat hyvin lyhyen aikavälin riippuvuuksissa, mutta pidemmän aikavälin riippuvuuksien opettaminen on hyvin epävakaa, ja tavalliset *RNN*-mallit eivät pysty oppimaan niitä helposti. Vastauksena tähän on kehitelty *LSTM (Long Short Term Memory)* mallit, jotka korjaavat opetuksen epävakauden pakottamalla opetusvaiheessa tasaisen

muutoksen verkon painojen ja vinoumien arvoihin. Oppimisen vakauden ansiosta LSTM-mallit kykenevät käsittelemään suuria datakokoelmia ja oppimaan datapisteiden välisiä suhteita, vaikka datapisteiden välillä olisi 1000 muuta datapistettä (Hochreiter ja Schmidhuber 1997, s. 1736)

Nimensä mukaisesti LSTM-malleissa on muistisoluja. Muistisolun koostuu solun tilasta sekä unohtamis-, tulo- ja lähtöporteista, jotka ohjaavat solun toimintaa. Muistisolun tila kuljettaa tietoa koko verkon läpi ja sen tila päivittyy jokaisella askeleella, unohtaen tarpeetonta tietoa ja tuoden uutta olennaista tietoa muistiin nykyisestä askeleesta. Unohtamisportti päättää, mikä osa nykyisen muistisolun tilasta tulisi unohtaa tai hylätä. Tuloportti määrittää, mitä uutta tietoa nykyisestä aika-askeleesta tulisi muistaa. Lähtöportti ohjaa mitä muistisolun tilasta välitetään eteenpäin seuraavalle aika-askeleelle tai ulostulo portille. (Hochreiter ja Schmidhuber 1997, s. 1744)

LSTM-mallit ovat olleet ensisijainen valinta aikasarjadataan käsittelyssä, mutta viime aikoina on kehitetty uusia malleja sarjadataan käsittelyyn, jotka saattavat olla tehokkaampia kuin LSTM-mallit.

## 2.3 Transformer-mallit ja huomio

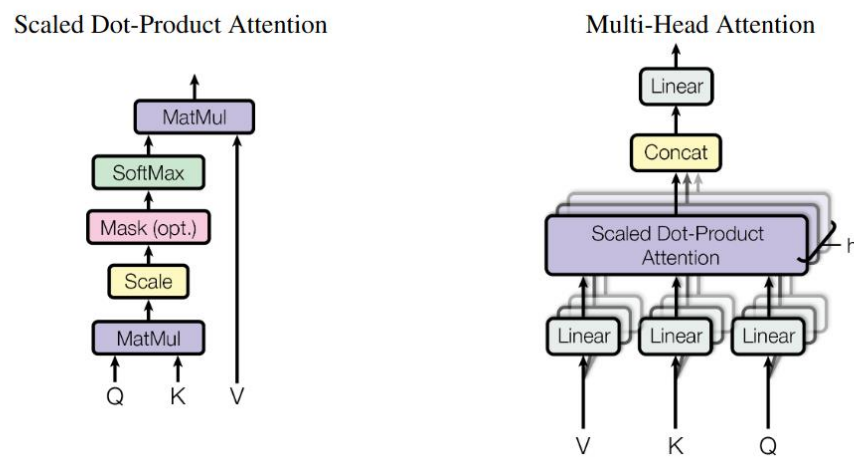
Transformerit pohjautuvat *enkooderi–dekooderi* arkkitehtuuriin, jossa enkooderikerrosten tarkoitus on muuntaa sisääntulo mallille ymmärrettäväksi jatkuvaksi representaatioksi, ja dekooderikerrokset luovat kyseisestä representaatiosta halutun ulostulon yksi osa kerrallaan. (Gomez & muut 2017, s. 2)

Transformer-mallien huomio mahdollistaa erikoistumisen monimutkaiseen *sekvenssida-taan*, kuten tekstipohjaiseen dataan. Yleiset käyttökohteet ovat esimerkiksi tekstin kääntäminen ja tekstin tuottaminen. Transformer-mallit ovat tulleet tunnetuiksi erityisesti GPT (Generative Pretrained Transformer) -mallien kautta. Vaikka Transformer-mallit kehitettiin tekstipohjaisen datan käsittelyyn, ne suoriutuvat hyvin myös muissa tehtävissä verrattuna perinteisempiin malleihin. (Szleski 2022, s. 323)

Transformer-mallien huomiota voidaan kuvata tietorakenteista tutulla dictionary tai map tietorakenteella, jossa on avain–arvo pareja. Avaimen kiinnitetty arvo saadaan ulos tietorakenteesta käyttämällä avainta. Huomion tapauksessa kaikki näistä on vektoreita. Meillä on kyselyvektori, joka kuvastaa huomion kohteen. Avainvektorilla saadaan kyselyvektorista arvovektorin. Arvovektorin avulla määritetään, kuinka paljon huomiota kyseinen osan kiinnittää muihin osiin nykyisellä haulla. (Gomez & muut 2017, s. 4)

Esimerkiksi jos Transformer-mallin huomiota vertaa tavallisiin lauseisiin, jokaisella sanalla on oma arvektori ja kyselyvektori. Jokaisen sanan kyselyvektori käytetään jokaisen sanan avainvektoriin, josta saadaan arvektorit jokaiselle sanalle. Arvektorit kertovat kuinka paljon sanojen merkitys riippuu kaikista muista sanoista.

Yhtä huomio kokonaisuutta, jossa on omat kyselyvektorit ja avainvektorit, kutsutaan *huomiopääksi*. Useita samanaikaisesti toimivia huomiokokonaisuuksia kutsutaan *monipäiseksi huomioksi (Multi-head attention)*. Monipäisen huomion tapauksessa malli voi kiinnittää huomiota moniin eri asioihin samaan aikaan. Kaikki huomioon liittyvät osat tulevat edellisen neurokerroksen ulostulosta aina seuraavaan kerrokseen, joten asiat joihin malli kiinnittää huomiota on täysin opittua opetusdatasta (Gomez & muut 2017, s. 5).



**Kuva 3. huomion matemaattinen toiminta (Gomez & muut 2017, s. 4)**

Kuva 3 näyttää, miten huomiota käsitellään Transformer-malleissa. Kuvassa Q (Query) vastaa hakua, K (Key) avainta ja V (Value) haun tuloksena saatua arvoa. Kuvasta nähdään myös, miten huomiopäät yhdistyvät monipäisessä huomiossa.

Transformer-malli vastaanottaa sekvenssin, joka kulkee usean enkooderikerroksen läpi. Jokainen kerros sisältää monipäisen huomion ja perinteisen neuroverkkotason. Enkooderissa jokainen osa pystyy kiinnittämään huomiota kaikkiin muihin osiin. Huomio kiinnittyy myös niihin osiin, jotka tulevat vasta kyseisen osan jälkeen. Enkoodauksen jälkeen syöte menee monen dekodauskerroksen läpi. Dekodauskerrokset ovat samanlaisia, mutta niissä on myös lisäelementti, joka kiinnittää monipäistä huomiota jo tuotettuun lopulliseen ulostuloon. Lisäksi dekooderin huomiomekanismi on rajoitettu niin, että se voi kiinnittää huomiota ainoastaan edeltäviin osiin. (Gomez & muut 2017, s. 3)

## 2.4 LSTM-verkot verrattuna Transformer-verkkoihin

Transformer-mallit toimivat täysin eri tavalla kuin RNN tai LSTM-mallit, sillä Transformer-malleissa ei ole muistisoluja. Transformer-mallien huomiopohjainen arkkitehtuuri mahdollistaa kaiken datan käsittelyn samanaikaisesti, ei aika-askelittain. (Szleski 2022, s. 323)

Vaikka LSTM-mallit on suunniteltu käsittelemään aikasarjadataa ja ne kykenevät oppimaan pitkävälisiä riippuvuuksia datasta, niiden yhtenä ongelmana säilyy aika-askelten yksittäinen läpikäynti. Jokaisella aika-askeleella LSTM-malli voi päättää unohtaa tärkeän osan nykyisestä muistista, mikä tekee pidempien kytkösten muodostamisen vaikeammaksi kuin lyhyempien. Koska Transformer-mallit näkevät kaiken datan samanaikaisesti, kytkösten aikavälin suuruus ei vaikuta niiden kykyyn oppia niitä. (Gomez & muut 2017, s. 6)

Transformer-mallien huomiokerrosten laskennallinen vaativuus riippuu datasta, mutta se on yleensä pienempi kuin RNN- ja LSTM-mallien rekursiivisissa kerroksissa. Lisäksi niiden laskenta voidaan suorittaa rinnakkain, toisin kuin RNN- tai LSTM-pohjaisissa malleissa. Nämä mahdollistavat suurempien mallien opettamisen samalla laskentateholla ja samassa ajassa. (Gomez & muut 2017, s. 6)

## 3. OSAKEMARKKINOIDEN ENNUSTAMINEN

Tässä luvussa kerrotaan hieman yleisemmin osakemarkkinoiden ennustamisesta, ennustamisessa käytetystä datasta, ennustamisen kohteista sekä ennustamisessa käytettyjen mallien menestyksen vertaamisesta

### 3.1 Osakemarkkinoiden data

Jotta tekoälymalleja voidaan käyttää osakemarkkinoilla, ne tarvitsevat dataa oppiakseen. Osakemarkkinoilla käytettävää dataa on paljon ja sitä pystytään keräämään monista eri lähteistä. Yleisimmin käytettyä dataa ovat osakkeiden kurssit, jotka kertovat osakkeiden hinnat kaikkina ajanhetkinä, jopa kymmeniä vuosia taaksepäin. Lisäksi voidaan käsitellä luovempia tiedonlähteitä, kuten osakkeen hinnasta johdetut signaalit, kyseisen yhtiön taloustiedot, osakekohtaisten uutisten optimistisuus, sosiaalisen median kommenttien odotukset osakkeesta ja laajemmat taloudelliset trendit.

### 3.2 Neuroverkkojen käyttökohteet

Osakemarkkinoilla ennustavia malleja voidaan käyttää monissa tarkoituksissa. Zhao ja muut (2023) sekä Li ja muut (2023) pyrkivät ennustamaan osakkeiden hinnan tietyn aikavälin päästä. Näin voidaan ennustaa, milloin tiettyä osaketta kannattaa ostaa ja milloin sitä kannattaa myydä, mutta paljon muitakin ennustamisen kohteita on olemassa. Ma, Wang ja Chen (2023) pyrkivät tutkimuksessaan ennustamaan kokoelman yhtiöitä, joista voidaan koota kaikkein tuottavin ja vakain osakesalkku. Tällöin meidän ei tarvitse huolehtia aktiivisesta osakkeiden ostamisesta ja myynnistä, vaan mallin menestys tulee suuremmista osakemarkkinoiden trendeistä. Liu ja muut (2024) käyttävät malliaan aktiivisesti S&P 500-indeksiin ostamiseen ja myymiseen ilman erillistä hinnan ennustamista.

### 3.3 Ennustustarkkuuden mittaaminen

Jotta mallien ennustamistuloksia pystyttäisiin vertaamaan, tarvitaan vertailuasteikko ja tapa mitata mallien menestystä. Yleinen vertailukohde on yksinkertaiseen osakkeen *ostamiseen ja pitämiseen* (Buy & Hold) vertailu (Zhao & muut 2023, s. 17). Tämä on hyvä perusmittakaava takaamaan, että malli on parempi kuin passiivinen odottaminen. Malleja voidaan verrata myös perinteisempiin malleihin, kuten lineaariseen regressioon ja satunnaismetsiin pohjautuviin malleihin.

Muita mittaustyökaluja on myös olemassa, jotka antavat paremman käsityksen mallin tuottavuudesta sekä luotettavuudesta. Hyvin yleinen tapa mitata mallien tarkkuutta liukulukujen ennustamisessa on *keskineliövirhe* (Mean Squared Error, MSE). Keskineliövirheen matemaattista kaavaa kuvaa yhtälö (1)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1)$$

missä  $n$  on ennustusten määrä,  $y$  on todellinen arvo ja  $\hat{y}$  on mallin antama ennuste.

Yleensä keskineliövirhe ei kuitenkaan aina anna tarpeeksi laajaa kuvaa mallin ennustamisesta, joten muitakin kaavoja käytetään. Koska keskineliövirhe nostaa virheen toiseen potenssiin, virheen suuruus vääristyy. Tähän vastauksena on *keskimääräinen absoluuttinen prosenttinvirhe* (Mean Absolute Percentage Error, MAPE), joka normalisoi virheen suuruuden riippumatta alkuperäisen luvun suuruudesta (Li & muut 2023 s.9). MAPE saadaan yhtälöstä (2)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

Näillä virhekaavoilla saadaan kuvailtua virhettä liukulukujen ennustamisessa, mutta osakemarkkinoiden ennustamisessa käytettyjä malleja voidaan arvioida muillakin tavoin. Mallien menestyminen on loppujen lopuksi kiinni niiden tuottavuudesta osakemarkkinoilla. Tuottavuutta voidaan mitata *kertyneellä tuotolla* (Accumulative Return, AR), joka ilmaisee saadun arvon verrattuna alkuperäiseen sijoitukseen (Liu & muut 2024, s. 22). Tuottavuuden kaava on (3)

$$AR = \frac{p_t}{p_0}, \quad (3)$$

missä  $p_t$  on nykyinen sijoituksen arvo ja  $p_0$  on alkuperäinen sijoitus

Kertyneestä tuotosta nähdään vain positiiviset tulokset, mutta mallin toiminnasta saadaan parempi kuva, kun tarkastellaan, kuinka paljon mallin tuottavuus muuttuu testijakson aikana. Liu ja muut (2024) selvittävät suurimman potentiaalisesti menetetyn voiton käyttämällä *suurimman arvonlaskun* (Maximum Drawdown, MDD) kaavaa. Suurin arvot lasku saadaan kaavasta (4)

$$MDD = \frac{p_{\text{häviö}} - p_{\text{huippu}}}{p_{\text{huippu}}}, \quad (4)$$

missä  $p_{\text{huippu}}$  on mikä tahansa huippu vertailuarvo ja  $p_{\text{häviö}}$  on pienin osakkeen arvo, joka on tullut vertailuarvon jälkeen. Mitä pienempi mallin MDD on, sitä paremmin malli kykenee välttämään saadun tuoton menetystä.

## 4. MALLIEN VERTAILUTULOKSET

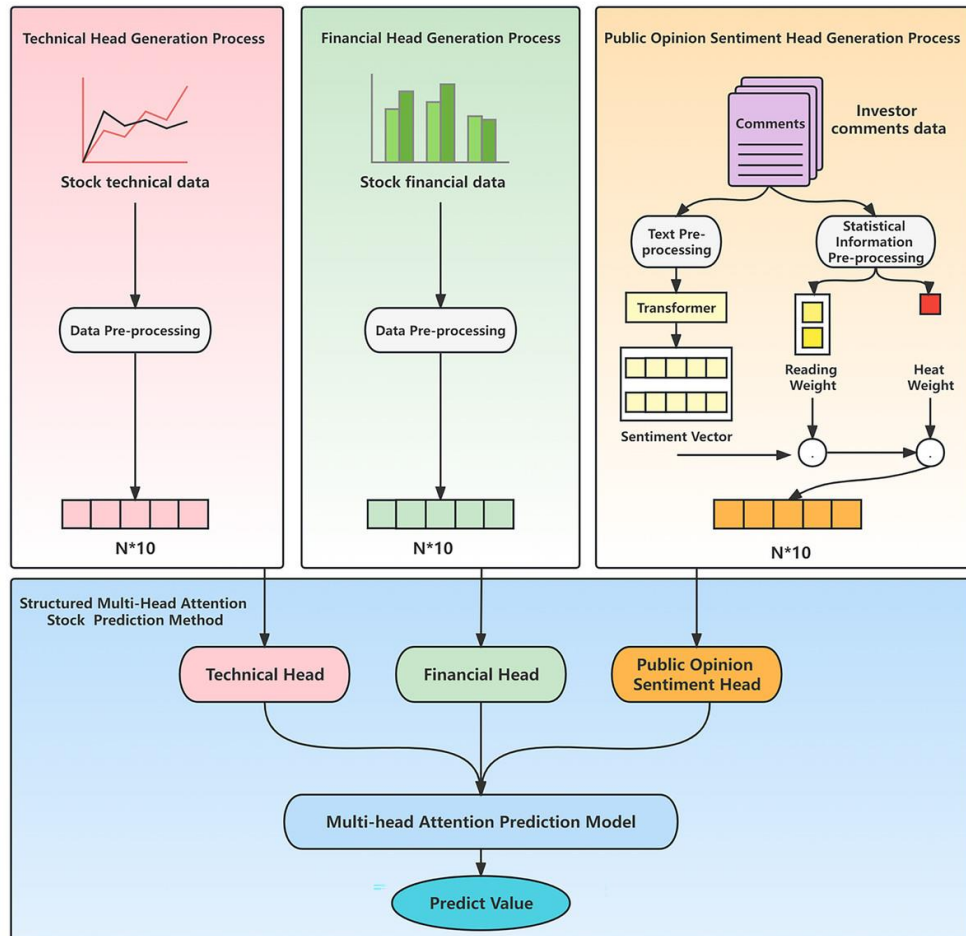
Tässä luvussa esitellään tutkimuksissa käytettyjen mallien erikoispiirteet sekä niiden opettamiseen ja testaamiseen liittyvät yksityiskohdat. Sen jälkeen tutkimusten saamia tuloksia pyritään normalisoimaan ja vertaamaan toisiinsa.

Joissain papereissa käytetään tavallisen LSTM-mallin lisäksi myös erikoisempia LSTM-malleja, mutta vertailussa huomioidaan ainoastaan tavalliset LSTM-mallit.

### 4.1 Käytettyjen mallien esittely

Ma, Wang ja Chen (2023) käyttivät Transformer- ja LSTM-malleja ennustamaan kokoelman yhtiöitä, joiden osakkeista saataisiin koottua kaikista tuottavin ja vakain osakesalkku. He käyttivät 80 eri indikaattoria, jotka koostuvat yhtiöiden taloudellisesta datasta ja muista makrotekijöistä. Kaksi identtistä mallia opetettiin samalla datalla. Toinen malli opetettiin ennustamaan osakesalkun tuottavuutta ja toinen taas sen epävakautta. Tasa-  
puolisuuden varmistamiseksi koottujen yhtiöiden osakesalkkujen arvioitu kokonaishinta pidettiin aina samana, jotta suurempien markkinatekijöiden vaikutus minimoitaisiin. Kokonaisuudessaan käsiteltyä dataa oli vuosilta 2000–2019 ja se jaettiin kolmeen osaan. Aluksi datasta ensimmäinen seitsemän vuotta oli opetusdataa, seuraava kolme vuotta validaatiodataa ja kaikki loput testidataa. Jokaisen testidatan testaamisen jälkeen, mallin opetusdataan lisättiin yksi vuosi lisää ja validaatiodataa vieritettiin vuoden eteenpäin ja malli opetettiin uudemmalle datalle ja testattiin uudelleen. Tätä jatkettiin, kunnes validaatiodata sisälsi kaikista uusimman datan.

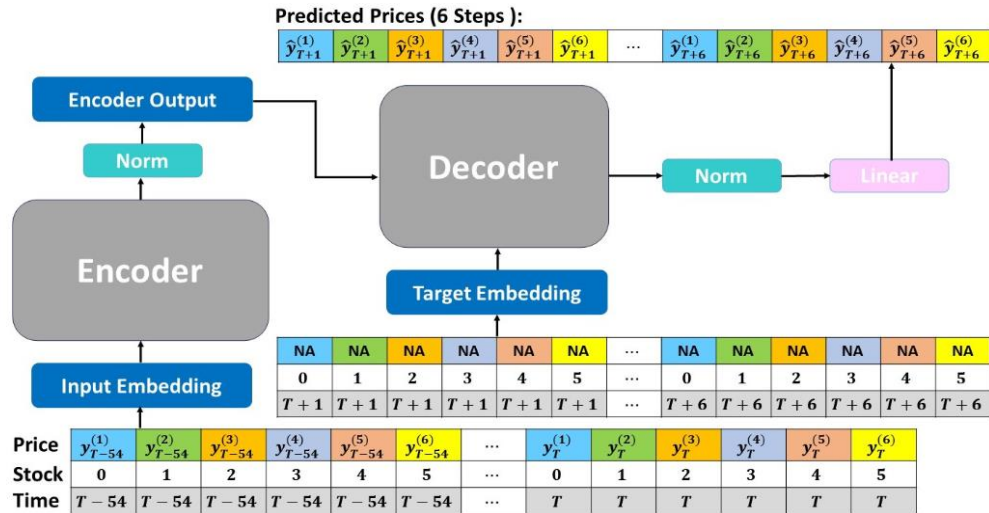
Zhao ja muut (2023) käyttivät Transformer-mallia, jossa oli kolme eri Transformer-huomiopäätä. Jokainen huomiopää käsitteli yhtä osaa datasta ilman, että ne näkivät muiden dataa. Yksi huomiopää käsitteli dataa, joka saatiin investointifoorumin kommenttien optimistisuudesta ja aktiivisuudesta, toinen keskittyi yhtiön taloudellisiin tilastoihin ja kolmas osakkeen teknisiin indikaattoreihin. Lopuksi malli yhdisti nämä kaikki huomiopäät ja täysin kytketyn neuroverkkokerroksen läpi malli tuotti ennustuksen osakkeen hinnasta.



**Kuva 4. kuvaus datan käsittelystä ja yhdistämisestä malliin (Zhao & muut 2023, s. 6)**

Kuva 4 esittelee yleisesti, millaista dataa huomiopäät käsittelivät, miten huomiopäät yhdistetään ja miten niiden avulla saadaan tuotettua ennustusarvo. Malli opetettiin käyttäen kerrallaan 12 kuukauden datamäärää, ja sitä testattiin seuraavan kuukauden opetusdatalla. Opetusdata alkoi tammikuusta 2019 ja jatkui joulukuuhun 2019, minkä jälkeen mallia testattiin vuoden 2020 tammikuun datalla. Seuraava opetusajankohta alkoi 2019 helmikuusta. Opettamista ja testaamista jatkettiin, kunnes opetusdata loppui vuoden 2022 toukokuuhun ja mallia testattiin vuoden 2022 kesäkuun datalla. Mallin opetettiin ennustaa osakkeiden sulkemishinta.

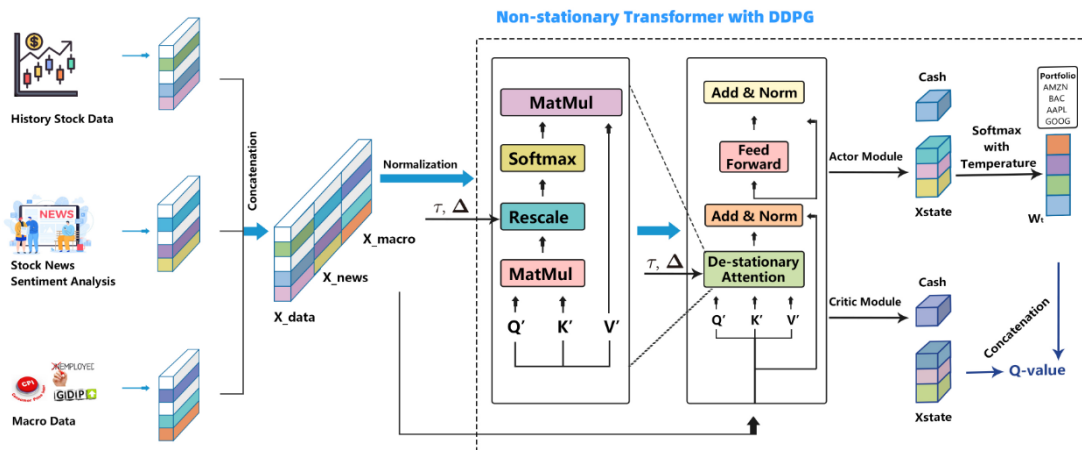
Li ja muut (2023) vertasivat tavallista LSTM-mallia sekä tavallista Transformer-mallia heidän kehittämänsä "Spacetimeformer"-nimiseen malliin. Spacetimeformer-malli kehitettiin käsittelemään dataa ajallisesti ja alueellisesti, eli toisiinsa läheiset datapisteet vaikuttavat toisiinsa enemmän. Mallin alueellisen komponentin tarkoitus on auttaa mallia paremmin huomioimaan, miten osakkeiden arvojen liikkeet vaikuttivat toisiinsa. Mallit käyttivät datanaan samanaikaisesti kuuden eri osakkeen hintadataa. Kyseiset osakkeet olivat TSMC, UMC, DELTA, EMC, FCFC ja YFH.



**Kuva 5. mallin rakenne ja opetusdata (Li & muut 2023, s. 9)**

Kuva 5 kuvastaa paremmin mallin rakennetta sekä näyttää miten malli vastaanottaa dataa ja tuottaa ennustukset. Hintadata oli jaettu 10 minuutin osiin, ja malli opetettiin käyttäen kerrallaan 55 osaa, mikä tarkoittaa, että jokaiselle kuudelle osakkeelle oli yhteensä hieman yli yhdeksän tunnin hintadataa per opetusaskel. Mallin tarkoituksena oli ennustaa opetusdatan perusteella seuraavien kuuden aika-askeleen hinta jokaiselle kuudelle osakkeelle samanaikaisesti. Mallin opetusdata kattoi ajanjakson kesäkuusta marraskuuhun 2022. Mallia testattiin ainoastaan kymmenen päivän datalla, 7.11.2022-18.11.2022 välisenä aikana.

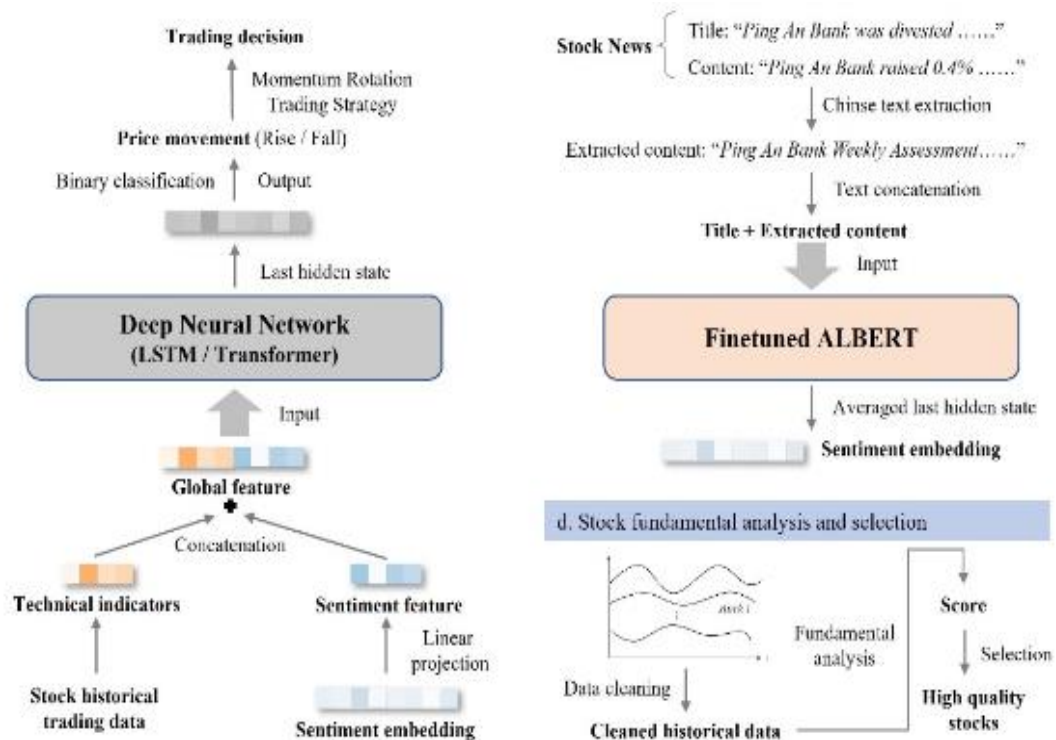
Liu ja muut (2024) vertailivat perinteistä LSTM-mallia, tavallista Transformer-mallia sekä heidän kehittämänsä 'Non-Stationary Transformer' (NSTD) -mallia. Heidän kehittämänsä malli, joka pohjautuu vahvistusoppimiseen, pyrkii ennustamaan suoraan optimaaliset hetket S&P 500-indeksin ostamiseen ja myymiseen, ilman että se ennustaa hintaa erikseen.



**Kuva 6. yleinen mallin toimintarakenne (Liu & muut 2024, s. 3)**

Kuva 6 esittelee paremmin käytetyn mallin datamuotoa sekä toimintaperiaatetta. Mallit käyttävät datanaan S&P 500-hintadataa, osakekohtaisten uutisten optimistisuutta ja taloudellisia makrotekijöitä, kuten bruttokansantuotetta, kuluttajahintaindeksiä ja työttömyysprosenttia. Malli hyödynsi vuosien 2012–2014 dataa oppimiseen, vuoden 2015 dataa validointiin ja vuoden 2016 dataa testaamiseen. Mallin tavoitteena oli analysoida viimeisten 10 päivän dataa ja oppia siitä, milloin on optimaalisinta ostaa tai myydä S&P 500-indeksiä.

Li, Hu ja Luo (2023) käyttivät LSTM- ja Transformer-mallia ennustamaan, tuleeko osakkeen hinta laskemaan vai nousemaan kahden päivän päästä ennustuspäivästä. Mallin opettamisessa käytettiin osakkeiden hintadataa, teknisiä indikaattoreita sekä osakekohtaisten uutisten ja uutisotsikoiden optimistisuutta. Mallin opetusdata kattoi ajanjakson 1.1.2020-30.6.2022, ja testidata oli ajalta 1.7.2022-31.12.2022.



**Kuva 7. optimistisuuden saaminen uutisista ja ulostuloarvon saaminen datasta (Li, Hu & Luo 2023, s. 3)**

Kuva 7 näyttää, miten uutisten optimistisuus saatiin ALBERT Transformerin avulla ja miten se yhdistetään teknisten indikaattorien kanssa, ennen kuin yhdistetty data syötetään malliin.

Mallia käytettiin yhdessä sijoitusstrategian kanssa, jossa valittiin 50 osaketta, jotka olivat osoittaneet suurinta ennustustarkkuutta mallin testausjakson aikana. Valituista osakkeista otettiin osakkeet, joiden sulkemishinta oli noussut kaikista eniten viimeisen 20 päivän aikana. Mallia käytettiin ennustamaan, tuleeko näiden osakkeiden hinta nousemaan

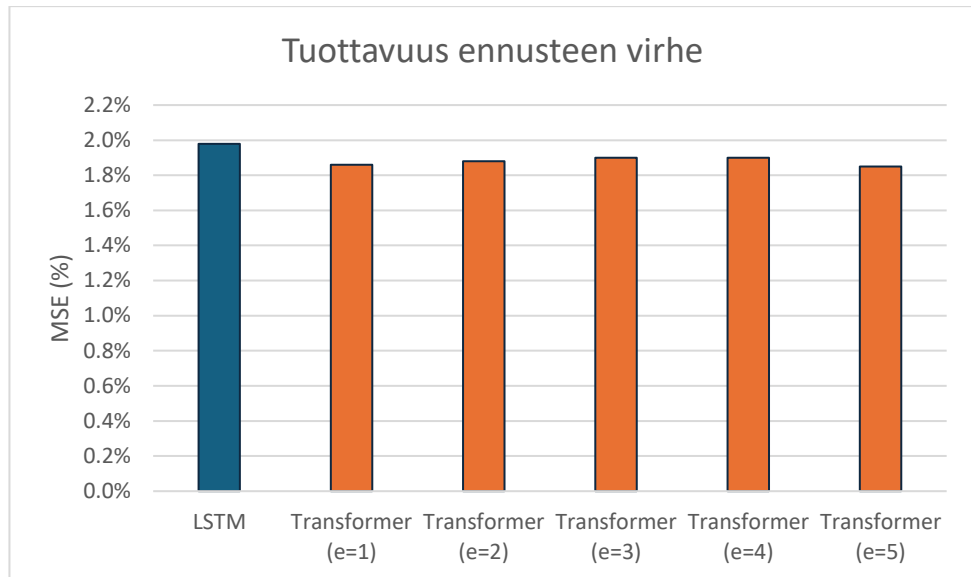
seuraavan kahden päivän päästä. Jos osakkeen hinta ennustettiin nousevan, osaketta ostettiin. Jos osaketta omistettiin ja malli ennusti sen hinnan laskevan seuraavan kahden päivän päästä, osake myytiin. Ostetut osakkeet myytiin myös automaattisesti, jos niiden hinta nousi yhteensä yli 22 % tai laski yli 8 %.

Wang ja muut (2022) pyrkivät mallilla ennustamaan CSI 300, S&P 500, N225 ja HSI indeksien seuraavan päivän sulkemishintaa. Mallien opetusdatana käytettiin ainoastaan normalisoituja indeksien päivittäisiä sulkemishintoja. Paperissa käytetty data saatiin väliltä 1.1.2010-31.12.2020. Kaikesta datasta alkava 80 % käytettiin mallien opettamisessa ja loput 20 % käytettiin mallien testaamiseen. Tutkimuksessa ei anneta tarkkaa opetus- ja testidatan alkamisen ja loppumisen ajankohtia, joten arvioitu opetusdata oli väliltä 1.1.2010-20.8.2018 ja arvioitu testidata oli väliltä 21.8.2018-31.12.2020. Tutkimuksen tulokset mallin tuottamasta tulosta sekä potentiaalisesta menetyksestä saatiin käyttämällä yksinkertaista sijoitusstrategiaa, jossa malli osti indeksiä, kun sen sulkemishinnan ennustettiin nousevan ja myi lyhyeksi indeksiä, kun sen sulkemishinnan ennustettiin laskevan.

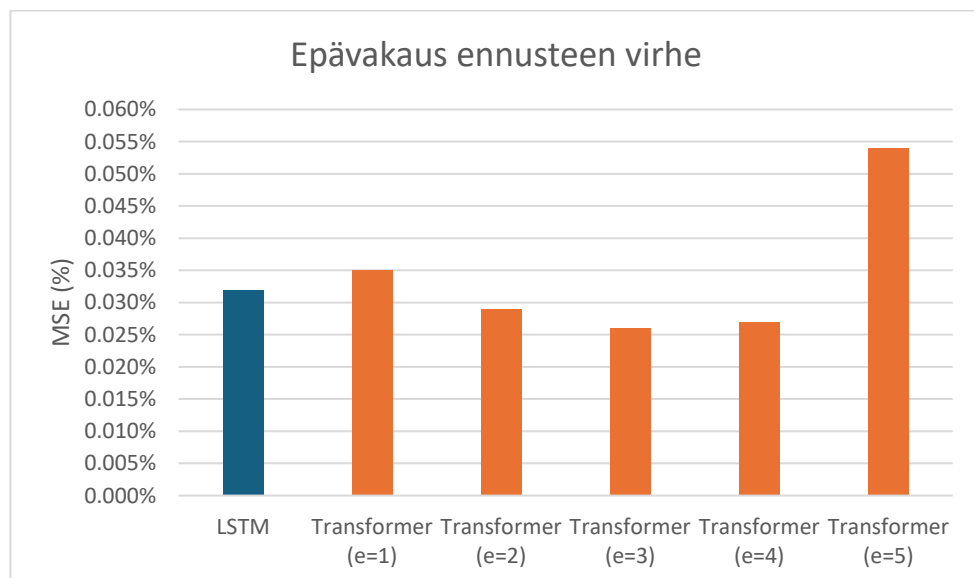
Mian (2023) käytti sekä LSTM- että Transformer-malleja ennustamaan osakkeiden sulkemishintoja. Mallit käyttivät datanaan osakkeiden teknisiä indikaattoreita, kuten päivittäinen ylin ja alin hinta, avaus- ja sulkemishinta, kohina sekä osakkeen kaupankäyntivolyymi. Tutkimuksessa käytetty data oli aikaväliltä 1.1.2017-17.9.2022, josta 70 % käytettiin mallien opettamiseen ja loput 30 % käytettiin mallien testauksessa. Opetus- ja testidatan alkamis- ja loppumisajankohtaa ei suoraan kerrota tutkimuksessa, mutta arvioitu opetusdata on väliltä 1.1.2017-31.12.2020 ja arvioitu testidata on väliltä 1.1.2021-17.9.2022.

## 4.2 Mallien tulosten vertailu

Ma, Wang ja Chen (2023) käyttivät LSTM-mallin lisäksi viittä eri Transformer-mallia, joissa oli yhdestä viiteen enkoodauskerrosta. Kuvissa 8, 9, 10 ja 11 kirjain "e" kuvastaa enkoodauskerrosten määrää.

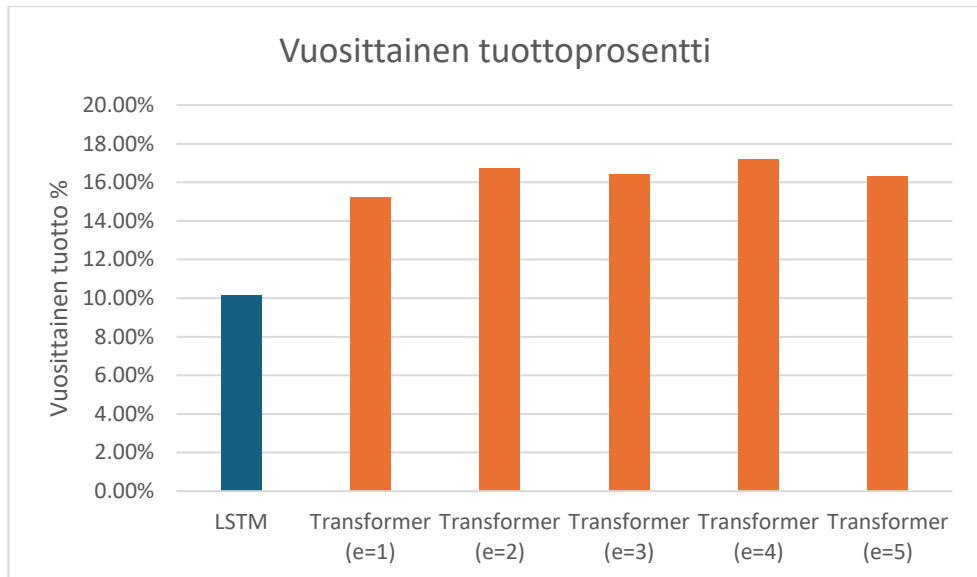


**Kuva 8. tuottavuus ennusteen virheen vertailu mallien välillä (Ma, Wang ja Chen 2023)**



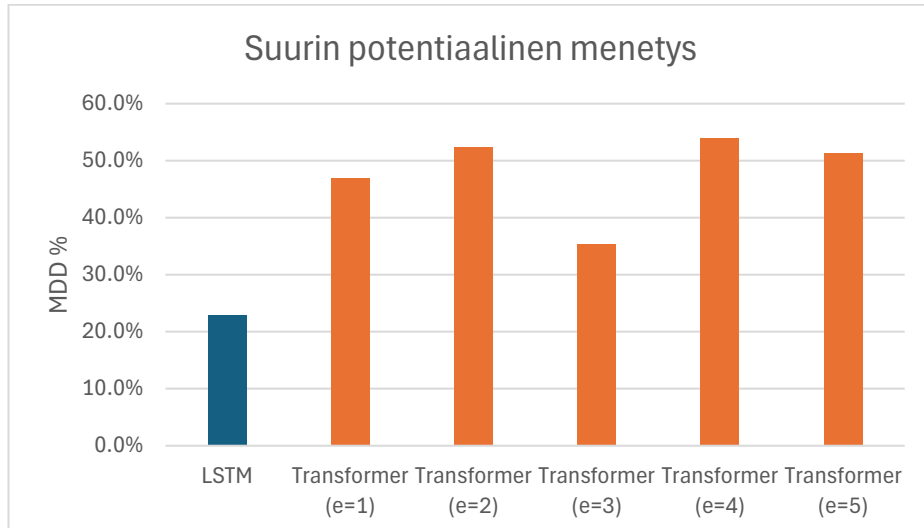
**Kuva 9. epävakaus ennusteen virheen vertailu mallien välillä (Ma, Wang ja Chen 2023)**

Tutkimuksen tuloksista koostetussa kuvan 8 näyttää tuottavuuden MSE ennustusvirhe on pienempi kaikilla Transformer-malleilla verrattuna LSTM-malliin. Kuitenkin kahdessa tapauksessa kuvassa 9 Transformer-mallin epävakauden ennustamisessa on suurempi virhe kuin verrattavalla LSTM-mallilla. Kuvista nähdään myös, että Transformerilla ei ole optimaalista määrää enkooderikerroksia. Eri määrä kerroksia oli parempi tuottavuuden ennustamisessa, kuin epävakauden ennustamisessa.



**Kuva 10. mallien vuosittainen tuotto prosentti (Ma, Wang ja Chen 2023)**

Kuva 10 esittää tutkimuksen tuloksista saatua LSTM-mallin ja Transformer-mallien ennustaman portfolioiden tuomaa vuosittaista tuottoa. Kuten tuloksista nähdään, kaikki Transformer-mallit kykenivät suurempaan vuosittaiseen tuotto prosenttiin kuin LSTM-malli.

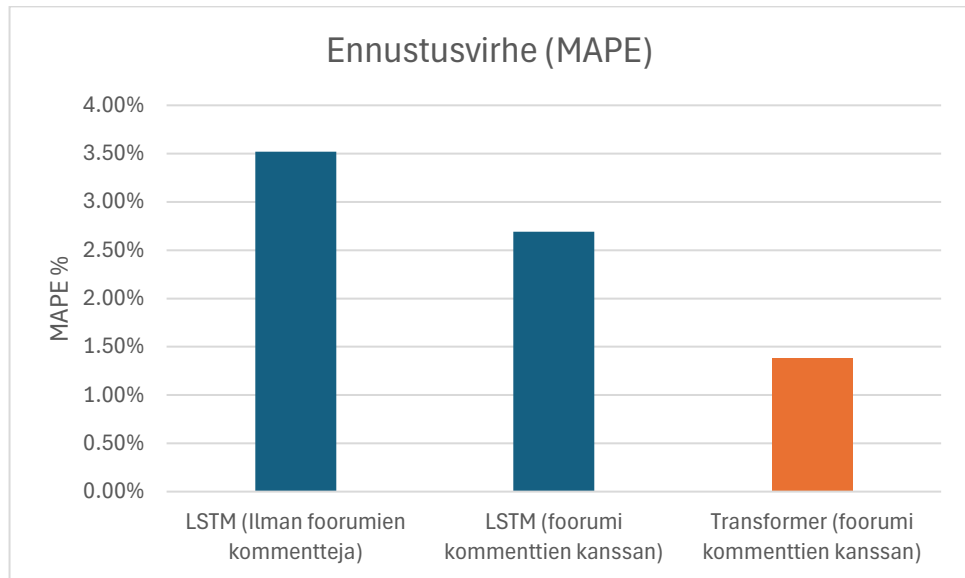


**Kuva 11. Suurin potentiaalinen menetys (Ma, Wang ja Chen 2023)**

Transformer-mallien valitsemien portfolioiden tuottavuus on suurempi, kuin LSTM-malleilla, mutta tuloksista koostetussa kuvassa 11 nähdään Transformer-mallien suosivan portfolioita, joiden tuottavuus on myös epävakaampaa.

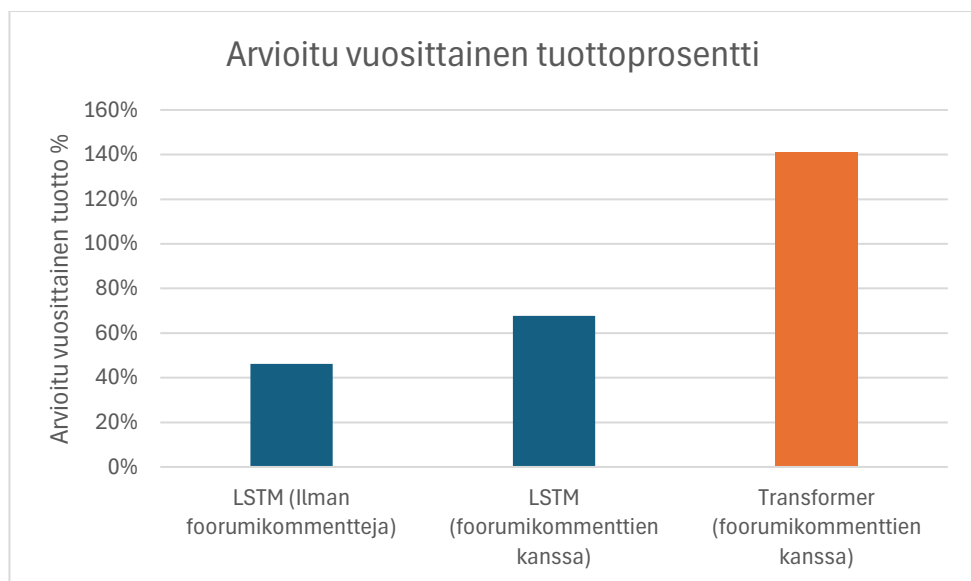
Zhao ja muut (2023) saivat samantapaisia tuloksia LSTM-mallien ja Transformer-mallin vertailusta. Transformer-mallien ennustusvirhe oli pienempi ja saatu tuotto prosentti oli

huomattavasti suurempi. On myös huomattava, että investointifoorumin kommenttien huomioon ottaminen pienentää ennustuksen virhettä huomattavasti.



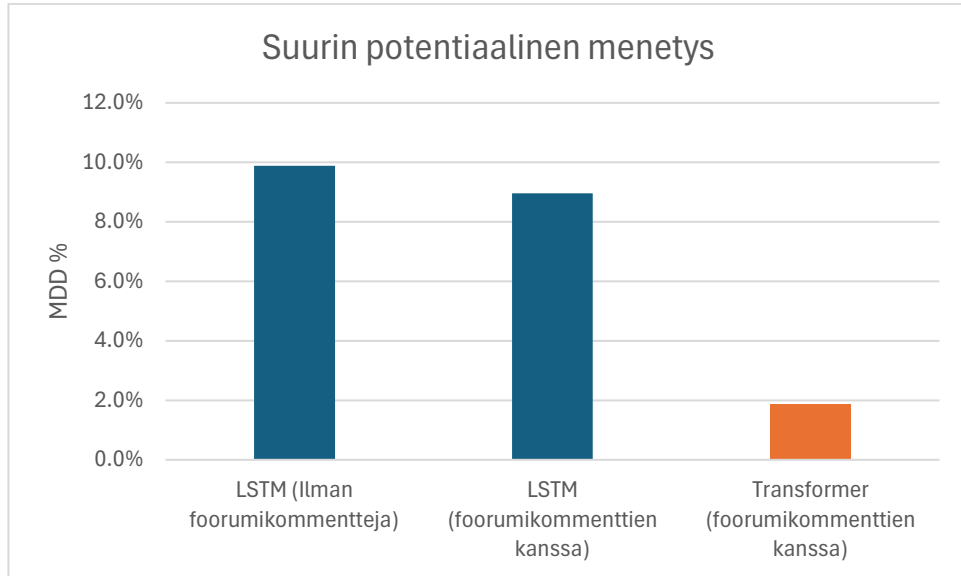
**Kuva 12. Ennustus virheen MAPE (Zhao & muut 2023)**

Kuvassa 12 on laskettu tutkimuksessa testattujen osakkeiden arvon MAPE ennustusvirheiden keskiarvot.



**Kuva 13. Arvioitu vuosittainen tuotto prosentti (Zhao & muut 2023)**

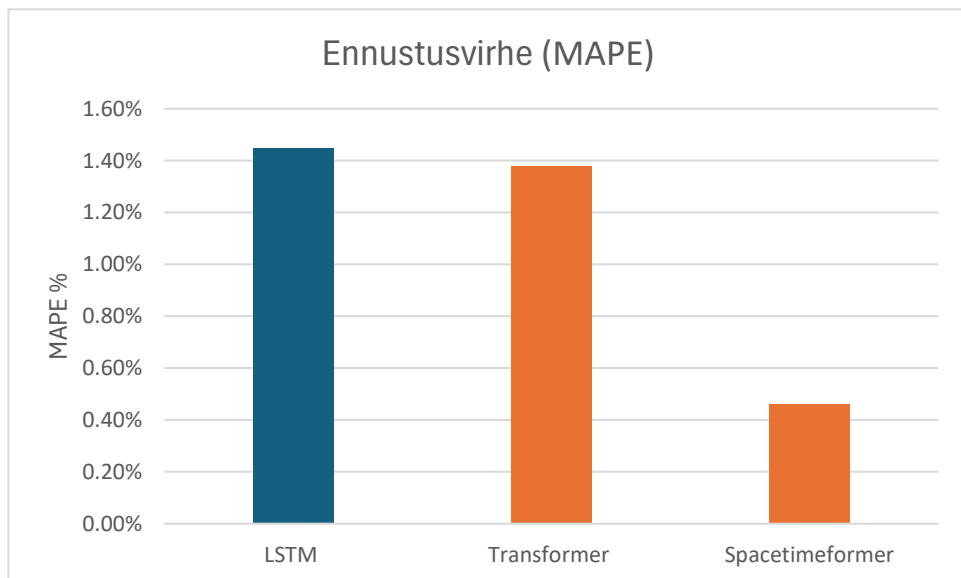
Kuvassa 13 näytetään tuloksien perusteella arvioitu vuosittainen tuotto prosentti. Vuosittainen tuotto prosentti on jouduttu arvioimaan, sillä tutkimuksessa malleja testattiin ainoastaan aikavälillä 20.10.2021-30.6.2022, joten kokonaiselle vuodelle ei riitä testidataa. Kyseiset arvioidut vuosituotto prosentti luvut on saatu kertomalla kaikkien tutkimuksessa testattujen osakkeiden tuotto prosenttiarvot murtoluvulla (365 / 254) ja ottamalla niistä keskiarvo.



**Kuva 14. Suurin potentiaalinen menetys (Zhao & muut 2023)**

Tutkimuksen tulosten mukaan Transformer-mallien tuottavuus oli myös vakaampaa, kuten kuvasta 14 nähdään.

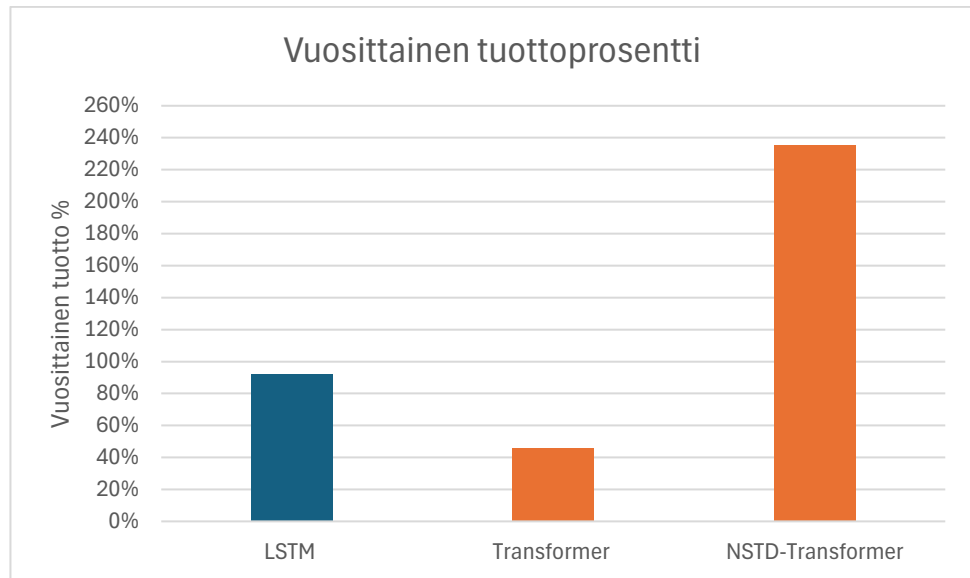
Li ja muut (2023) keskittyivät pääasiassa heidän itse kehittämään Spacetimeformer-malliin, mutta he käyttivät vertailussaan myös tavallista Transformer-mallia. Vertailussa Spacetimeformer-malli kykeni saamaan pienimmän ennustusvirheen, mutta myös tavallinen Transformer oli parempi kuin vertailtu LSTM-malli.



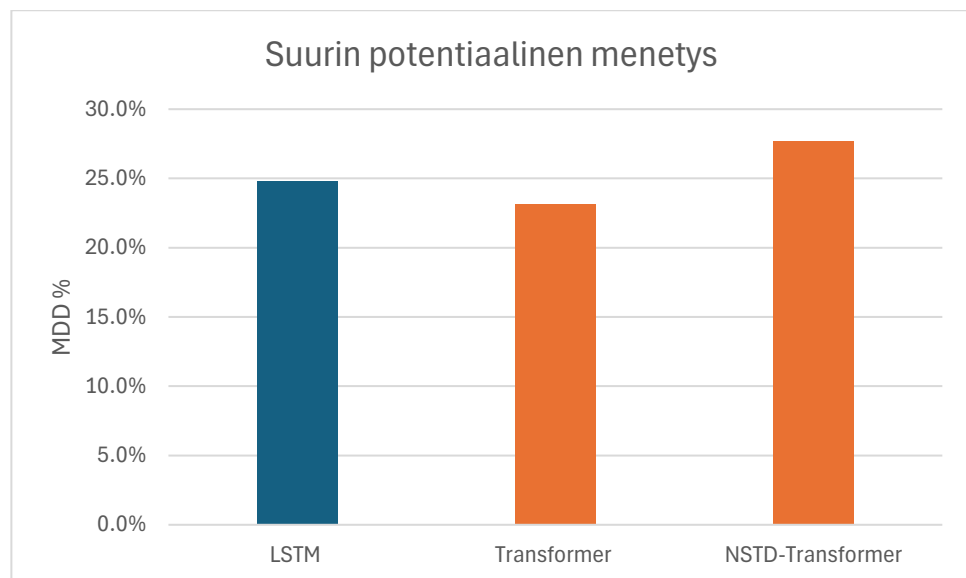
**Kuva 15. Ennustus virheen MAPE (Li & muut 2023)**

Tuloksista koostettu kuvan 15 MAPE arvot saatiin ottamalla keskiarvo kaikkien kuuden osakkeen keskiarvo MAPE arvoista.

Liu ja muut (2024) vertasivat heidän itse kehittämää NSTD-Transformeria, mutta vertailussa oli mukana myös tavallinen Transformer-malli sekä LSTM-malli. Tuloksista koostetusta kuvasta 16 nähdään, että heidän tavallinen Transformer implementaatio pärjasi huonommin kuin LSTM-malli, mutta NSTD-Transformer pärjasi kaikista parhaiten. Tuloksista luotu kuva 17 näyttää, miten tavallinen Transformer-mallin tuottavuus oli kuitenkin vakaampaa, kuin LSTM-malli tai NSTD-Transformer-malli.



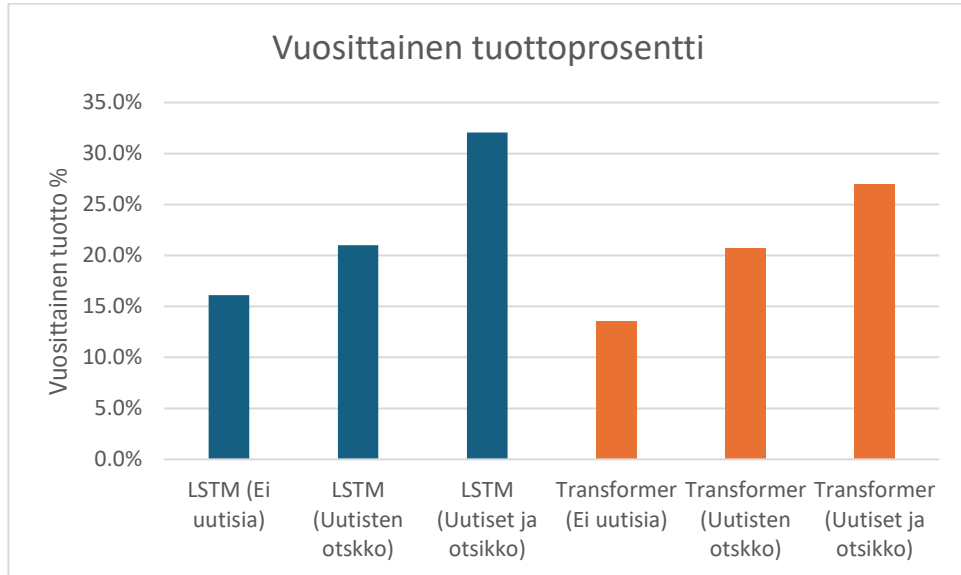
**Kuva 16. Vuositainen tuottoprosentti (Liu & muut 2024)**



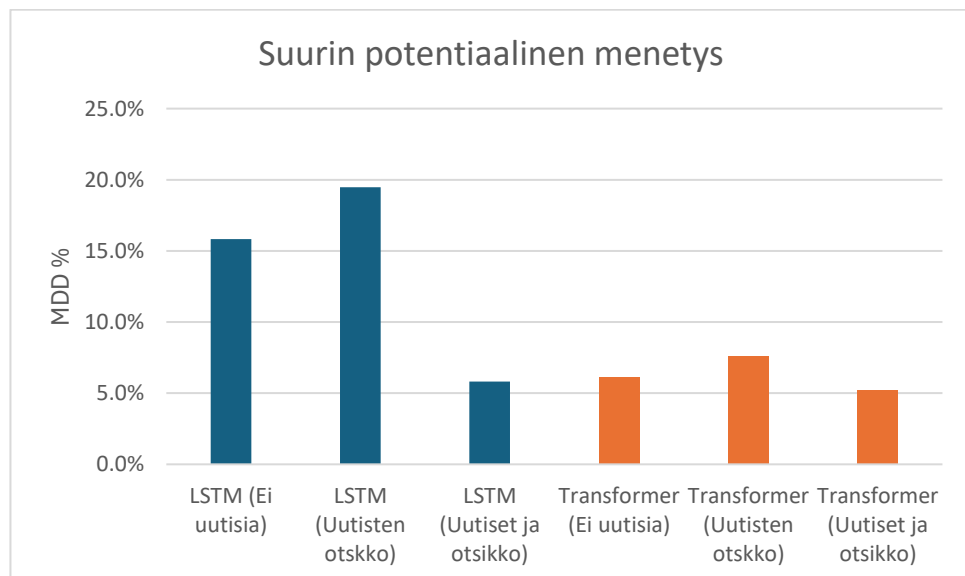
**Kuva 17. Mallien suurin potentiaalinen menetys (Liu & muut 2024)**

Li, Hu ja Luo (2023) tulokset olivat paljon tasaisempia. Tuloksista saatu kuva 18 kuvastaa miten heidän LSTM-mallinsa sai suuremman vuosittaisen tuoton kuin Transformer-malli,

sekä uutisdatan kanssa että ilman uutisdataa. Transformer-mallin saama tuotto oli kuitenkin vakaampaa kuin LSTM-mallilla. Kuten tutkimuksen tuloksista saadussa kuvassa 19 näemme, Transformer-mallin suurin potentiaalinen menetys oli aina pienempi kuin LSTM-mallilla samalla datalla.



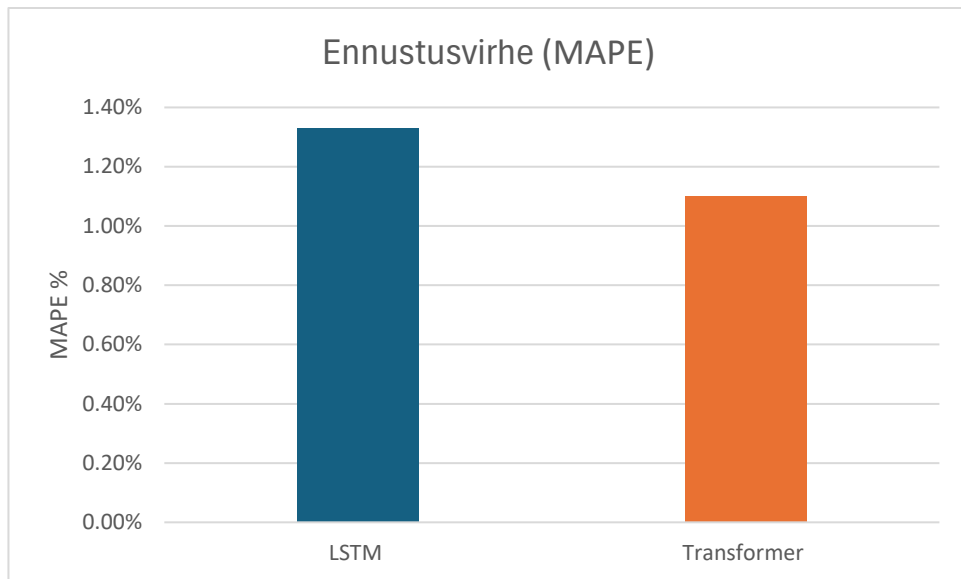
**Kuva 18. LSTM- ja Transformer-mallien vuosittainen tuottoprosentti (Li, Hu & Luo 2023, s.12)**



**Kuva 19. LSTM- ja Transformer-mallien suurin potentiaalinen häviö (Li, Hu & Luo 2023, s.12)**

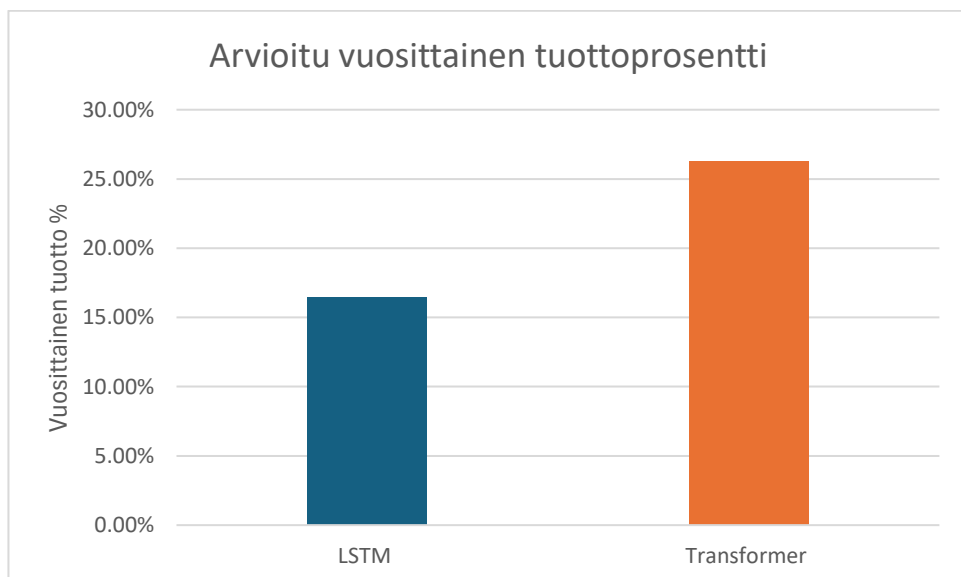
Kuvista myös näkee, kuinka paljon uutisdatan huomioon ottaminen auttoi malleja saamaan paljon suuremman tuottavuuden sekä LSTM-mallilla, että Transformer-mallilla.

Wang ja muut (2022) tuloksien mukaan Transformer-mallien ennustusvirhe oli pienempi kuin LSTM-malleilla. Tuloksista saatu kuva 20 kuvastaa miten Transformer-mallin ennustusvirhe vertaa LSTM-malliin



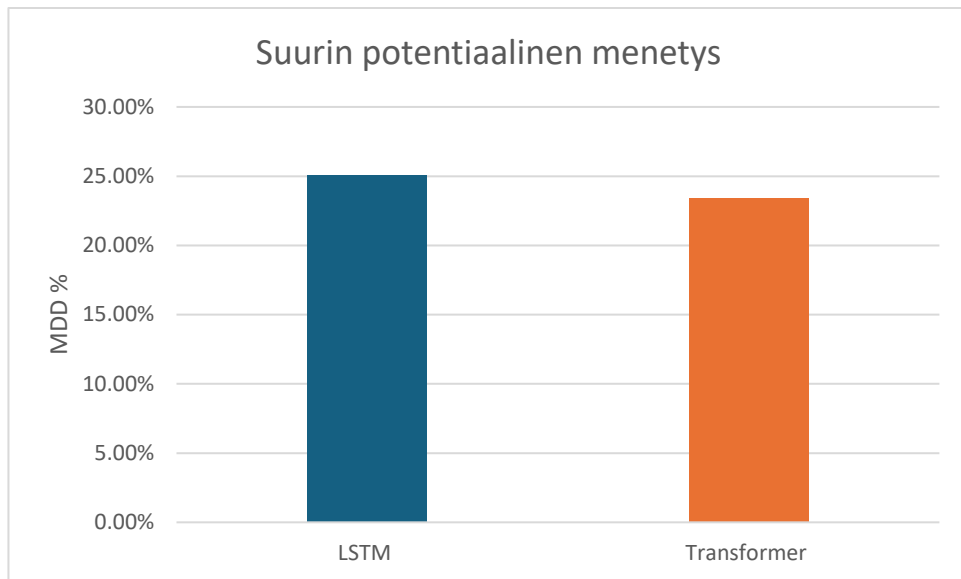
**Kuva 20. Transformer- ja LSTM-mallien ennustusvirhe (Wang & muut 2022)**

Tuloksissa Transformer-mallien tuotto oli suurempi jokaisella indeksillä verrattuna LSTM-malleihin. Kuva 21 näyttää arvioitun vuosittaisen tuoton, joka saatiin ottamalla keskiarvo tutkimuksessa raportoitujen indeksien kokonaistulosta ja jakamalla se murto-luvulla (803 / 365). Vuosittainen tuotto jouduttiin arvioimaan, sillä tutkimuksen raportoima kokonaistuotto oli koko testiajalta, jonka arvioitu kesto aika oli yli kaksi vuotta.



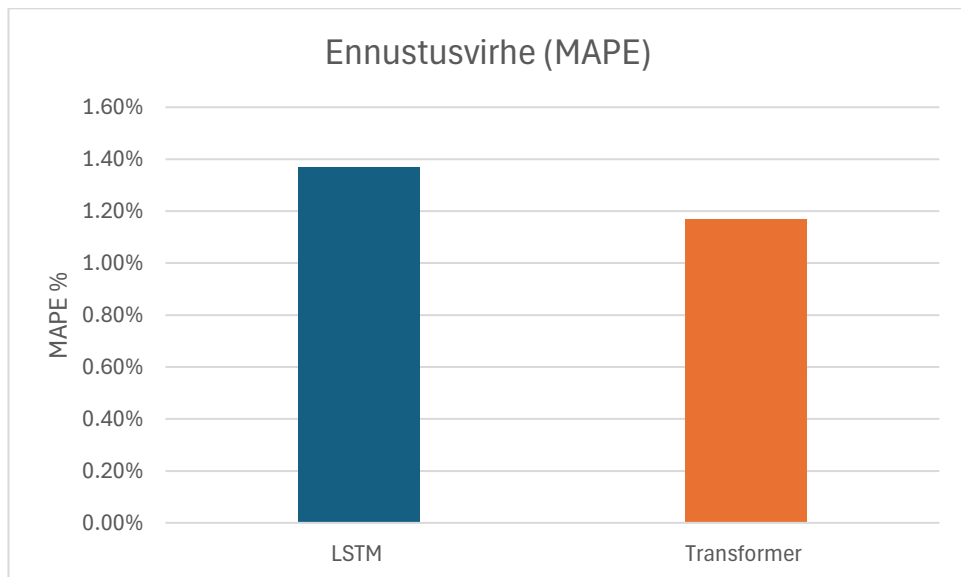
**Kuva 21. Transformer- ja LSTM-mallien arvioitu vuosittainen tuotto prosentti (Wang & muut 2022)**

Transformer-mallien tuotto oli myös vakaampaa testiajalla. Kuva 22 kuvastaa, miten LSTM-mallin suurin potentiaalinen menetys testiajalla oli keskiwertona Transformer-mallia suurempi.



**Kuva 22. Transformer- ja LSTM-mallien suurin potentiaalinen menetys (Wang & muut 2022)**

Mian (2023) tutkimuksen tuloksien mukaan Transformer-mallin ennustustarkkuus oli parempi kuin LSTM-mallilla. Kuva 23 on tutkimuksen tuloksista testiajalta saatujen ennustustarkkuuksien keskiarvo.



**Kuva 23. Transformer- ja LSTM-mallien ennustusvirhe (Mian 2023)**

## YHTEENVETO

Tutkimusten tulokset osoittavat, että sekä LSTM- että Transformer-mallit ovat tehokkaita työkaluja aikasarjadatan, kuten osakemarkkinadatan, ennustamisessa. Vaikka useissa tutkimuksissa Transformer-mallit suoriutuivat paremmin kuin LSTM-mallit, tulosten ja testiolosuhteiden vaihtelevuus vaikeuttaa lopullisten johtopäätösten tekemistä.

Useissa tutkimuksissa malleja opetettiin datalla ilman osakekohtaista optimistisuutta sekä datalla, joka sisälsi osakekohtaisten uutisten ja foorumikommenttien optimistisuuden. Näissä tutkimuksissa optimistisuus datan huomioiminen mallien opettamisessa sekä testaamisessa auttoi malleja suoriutumaan paremmin. Tästä huomataan, että ennustukseen käytetyn mallin valitseminen on vain yksi osa suurempaa kokonaisuutta ja suorituskykyyn vaikuttaa moni muukin tekijä.

Riippumatta siitä, minkä mallin valitsee datan ennustamiseen, Transformer-mallit ovat sopiva valinta saamaan optimistisuusdataa tekstipohjaisista lähteistä.

Jatkotutkimuksissa voitaisiin tarkastella muita uusia sarjadataan kehitettyjä malleja, kuten esimerkiksi vuonna 2023 esiteltyä MAMBA-mallia, joka on suunniteltu kilpailemaan Transformer-mallien kanssa. Lisäksi tutkimuskohteiksi voivat soveltua eri datatyypit ja datalähteet, sekä niiden vaikutus osakehinnan ennustusvirheeseen. Lisäksi tässä tutkimuksessa ei käsitelty yhdistelmä-malleja, jotka hyödyntävät sekä LSTM- että Transformer-mallien osia tai kerroksia.

# LÄHTEET

- Gomez, A. N. et al. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*. 30, pp. 11.
- Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*. 9(8), pp. 1735–1780.
- Li, Y.-C. et al. (2023). Stock Market Forecasting Based on Spatiotemporal Deep Learning. *Entropy (Basel, Switzerland)*. 25(9), 1326.
- Li, W., Hu, C. & Luo, Y. (2023). A Deep Learning Approach with Extensive Sentiment Analysis for Quantitative Investment. *Electronics (Basel)*. 12(18), 3960.
- Liu, Y. et al. (2024). Revolutionising Financial Portfolio Management: The Non-Stationary Transformer's Fusion of Macroeconomic Indicators and Sentiment Analysis in a Deep Reinforcement Learning Framework. *Applied Sciences*. 14(1), 274.
- Ma, T., Wang, W. & Chen, Y. (2023). Attention Is All You Need: An Interpretable Transformer-Based Asset Allocation Approach. *International Review of Financial Analysis*. 90, 102876.
- Mian, T. S. (2023). Evaluation of Stock Closing Prices Using Transformer Learning. *Engineering, Technology & Applied Science Research*. 13(5), pp. 11635–11642.
- Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*, 2<sup>nd</sup> ed.
- Wang, C. et al. (2022). Stock Market Index Prediction Using Deep Transformer Model. *Expert Systems with Applications*. 208, 118128.
- Zhao, C. et al. (2023). A Structured Multi-Head Attention Prediction Method Based on Heterogeneous Financial Data. *PeerJ Computer Science*. 9, e1653.