

Emma Sjöholm

SYVÄNEUROVERKKOJEN TARKKUUS OSAKEKURSSIEN ENNUSTAMISESSA

Kandidaatintutkielma
Informaatioteknologian ja viestinnän tiedekunta
Joulukuu 2022

TIIVISTELMÄ

Emma Sjöholm: Syväneuroverkkojen luotettavuus osakekurssien ennustamisessa
Kandidaatintutkielma
Tampereen yliopisto
Tietotekniikan tutkinto-ohjelma
Joulukuu 2022

Osakekurssien ennustaminen on tärkeää ja mielenkiintoista, mutta myös vaikeaa. Luotettavalla mallilla sijoittaja voi tehdä suurta voittoa ja minimoida sijoitustensa riskit. Luotettavan mallin kehittäminen onkin tärkeää niin yksilöille kuin yrityksillekin. Tutkijat kehittävätkin jatkuvasti uusia malleja ennustamaan osakekurssien kehitystä entistä tarkemmin. Osakkeiden hintoihin vaikuttavat kuitenkin useat muuttujat. Tällä vuosituhanella useat syväoppimismenetelmät ovat korvanneet perinteiset tilastolliset mallit.

Tämä tutkielma on toteutettu kirjallisuuskatkauskana. Tavoitteena on selvittää miten GRU- ja LSTM-neuroverkoilla ennustetaan osakekurssien kehitystä, miten tarkkoja osakekurssien kehityksen ennusteet ovat ja mitkä tekijät vaikuttavat tähän tarkkuuteen. Ensiksi perehdytään siihen, mitä osakekurssien ennustaminen oikeastaan on ja millaisilla mittareilla ennustuksen tarkkuutta mitataan. Tämän jälkeen tutustutaan neuroverkkojen ja syväoppimisen peruskäsitteistöön ja takaisinkytkettyihin neuroverkkoihin. Lopuksi käydään läpi miten takaisinkytkettyjä neuroverkkoja koulutetaan ennustamaan osakedataa ja millaisia tarkkuuksia eri tutkimuksissa ollaan saavutettu.

Tutkimus osoittaa, ettei LSTM:stä tai GRU:sta kumpikaan ylivoimainen, vaan suoritusteho vaihtelee datajoukkojen ja koulutustavan välillä. Objektiiivisen vertailun tekeminen ei kuitenkaan ole helppoa. Koulutuksessa on kiinnitettävä huomiota siihen, mitä neuroverkoille syötetään. Turha syötedata voi heikentää tuloksia, mutta toisaalta myös oleellisen syötedatan pois jättäminen saattaa vähentää tarkkuutta.

Avainsanat: osakekurssien ennustaminen, syväneuroverkot, takaisinkytketyt neuroverkot, LSTM, GRU

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

ABSTRACT

Accuracy of deep neural networks in stock price prediction
Bachelor's thesis
Tampere University
Information Technology
December 2022

Predicting stock prices is important and interesting, but also difficult. With an accurate model, an investor can make a big profit and minimize the risks of their investments. Developing an accurate model is important for individuals and businesses alike. Scientists are constantly developing new models to predict stock price trends even more accurately. However, several variables affect stock prices. In this millennium, several deep learning methods have replaced traditional statistical models.

The research has been conducted as a literary review. The goal of is to find out how the GRU and LSTM neural networks predict stock price trends, how accurate stock price forecasts are, and what factors influence this accuracy. First, some basic theory on price forecasting and prediction accuracy measurement is introduced. This is followed by familiarization with the basic concepts of neural networks and deep learning and feedback neural networks. Finally, we will look at how recurrent neural networks are trained to predict stock data and what accuracy has been achieved in various studies.

The study shows that neither LSTM nor GRU is superior, but throughput varies between data sets and training methods. However, making an objective comparison is not easy. In training, attention should be paid to what is being fed to neural networks. Pointless input data can reduce results, but on the other hand, the exclusion of essential input data.

Keywords: stock price prediction, deep neural networks, recurrent neural networks, LSTM, GRU

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

SISÄLLYSLUETTELO

1. JOHDANTO	1
2. OSAKEKURSSIEN ENNUSTAMINEN JA SUORITUSMITTARIT	3
2.1 Osakekurssien ennustaminen	3
2.2 Keskipoikkeama (MAE) ja prosentuaalisen virheen keskiarvo (MAPE)	4
2.3 Keskineliövirhe (MSE) ja keskineliövirheen neliöjuuri (RMSE)	5
2.4 Selitysaste	5
3. TAKAISINKYTKETYT NEUROVERKOT	7
3.1 Neuroverkot ja syväoppiminen	7
3.2 Takaisinkytketyt neuroverkot (RNN)	8
3.3 LSTM-neuroverkko	9
3.4 GRU-neuroverkko	10
4. SYVÄNEUROVERKOT OSAKEKURSSIEN ENNUSTAMISESSA	11
4.1 Käytetyt syväoppimismallit	11
4.2 Mallien tarkkuus ennustamisessa	15
YHTEENVETO	20
LÄHTEET	21

LYHENTEET JA MERKINNÄT

GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
MSE	engl. Mean Squared Error, keskineliövirhe
RMSE	engl. Root Mean Squared Error, keskineliövirheen neliöjuuri
MAE	engl. Mean Absolute Error, keskipoikkema
MAPE	engl. Mean Absolute Percentage Error, prosentuaalisen virheen keskiarvo

1. JOHDANTO

Tekoälystä on tullut keskeinen teknologia 2010-luvun aikana ja se on nykyään merkittävä työkalu muun muassa rahoituselämässä. Yksi keskeinen sovelluskohde tekoälylle rahoituselämässä on sijoitusten arvokehityksen ennustaminen. Sijoittajat ovat aina halunneet etsiä uusia ja uniikkeja tekniikoita voidakseen ennustaa onnistuneesti osakehintoja ja tehdä taloudellista voittoa ja vähentää päätöksenteon riskejä (Dey ja muut 2021, s. 1; Liu ja muut 2022, s. 1).

Osakehintojen ennustaminen on tärkeä tutkimuskohde, koska sillä on merkitystä niin yksilöille, yrityksille kuin hallituksillekin (Alkhatib ja muut 2022 s. 1). Datan epälineaarisuus ja -vakaus tekevät kuitenkin osakehintojen ennustamisesta erittäin vaikeaa (Patra & Mohanty 2022 s. 2). Osakehintoihin vaikuttavat tekijät ovat monimutkaisia ja muuttuvia, ja osakehintojen vaihtelut sisältävät myös jonkin verran satunnaisuutta (Lv ja muut 2022 s. 1). Nämä haasteet ovat syy siihen, minkä vuoksi koneoppimismenetelmien soveltaminen ennustamisessa kiinnostaa tutkijoita (Alkhatib ja muut 2022 s. 1-2). Nykypäivänä vakaan mallin kehittäminen on helpottunut tekoälyn kehityksen, suuren mittakaavan datan saatavuuden sekä kasvaneen laskennallisen kapasiteetin ansiosta (Pokthrel ja muut 2022 s. 2).

Eri tyyppisiä RNN:iä (recurrent neural network) eli takaisinkytkettyjä neuroverkkoja käytetään aikasarja-analyysin teossa, etenkin osakehintojen ennustamisessa (Dey ja muut 2021 s. 1). RNN:istä esimerkiksi LSTM- (Long Short-Term Memory) ja GRU-neuroverkot (Gated Recurrent Unit) ovat suosittuja syväoppimisarkkitehtuureja osakemarkkinoiden ennustamiseen (Shahi ja muut 2020 s. 2; Wang ja muut 2021 s. 1). Tämän tutkielman tavoitteena on avata, mitä ovat LSTM- ja GRU-neuroverkot ja verrata niiden luotettavuutta osakekurssien ennustamisessa.

Tämä tutkimus on toteutettu kirjallisuuskatsauksena. Haut on tehty Andorissa. Tietokantahaut toteutettiin käyttämällä hakulausekkeena "GRU" AND "LSTM" AND ("stock price prediction" OR "stock price forecasting"). Aineistojen julkaisuvuosi rajattiin aikavälille 2020-2022, jotta saataisiin viimeaikaista tutkimustietoa. Lisäksi rajattiin saatavuudeksi vertaisarvioidut lehdet, jotta aineisto olisi luotettavaa. Tällä ensimmäisellä rajauksella saatiin 16 tulosta. Toisessa rajauksessa valittiin 12 lähdettä

tiivistelmän ja yhteenvedon perusteella. Lähteet luettiin vielä kokonaan. Lisäksi otettiin kaksi tekoälyyn liittyvää lähdeä täydentämään teoriaa tutkimusten taustalla.

Tutkielma vastaa seuraaviin tutkimuskysymyksiin:

- Miten GRU- ja LSTM-neuroverkkoja koulutetaan ennustamaan osakekursseja?
- Miten ne eroavat luotettavuudeltaan osakekurssien ennustamisessa?
- Mitkä tekijät vaikuttavat ennustusten luotettavuuteen?

Työ tutkii takaisinkytkettyjen neuroverkkojen käyttöä osakekurssien kehityksen ennustamisessa. Luvussa 2 käsitellään osakekurssien ennustamiseen liittyviä käsitteitä ja taustaa. Luvussa 3 avataan syväoppimista sekä tutkielman keskeisimpiä syväneuroverkkoihin liittyviä käsitteitä. Luvussa 4 verrataan LSTM:n ja GRU:n luotettavuutta osakekurssien ennustamisessa. Luvussa 5 esitetään tutkielman yhteenvedo ja päätelmät.

2. OSAKEKURSSIEN ENNUSTAMINEN JA SUORITUSMITTARIT

Tässä luvussa käsitellään lyhyesti osakekurssien ennustamisen historiaa ja esitellään mittareita, joilla voidaan arvioida ennustusten tarkkuutta.

2.1 Osakekurssien ennustaminen

Tarkan ennusteen luominen osakedatalle ei ole helppoa johtuen datan epälineaarisesta ja -vakaasta luonteesta. On olemassa kaksi eri mieltä olevaa teoriaa, tehokkaiden markkinoiden hypoteesi (engl. efficient market hypothesis, EMH) ja satunnaiskulkuteoria (eng. random walk theory). Tehokkaiden markkinoiden hypoteesin mukaan tulevaisuuden osakedataa voidaan ennustaa historiallisen osakedatan pohjalta. Sen sijaan satunnaiskulkuteorian mukaan tuleva osakedata ei riipu historiallisesta osakedatasta eikä historiallisesta datasta voida löytää mitään käyttökelpoisia säännönmukaisuuksia, joiden avulla voitaisiin ennustaa tulevia osakejaksoja (Shahi ja muut 2020 s. 1).

Osakkeen hintaan vaikuttavat tekijät voidaan jakaa makro- ja mikro-tasoihin. Makrotason tekijöitä esimerkiksi bruttokansantuote (BKT), yleinen korkotaso sekä inflaatio. Makrotason tekijöiden välinen vuorovaikutus vaikuttaa talousindikaattoreihin. BKT:n kasvu tarkoittaa, että yritysten liikevaihdot, tulokset ja myös osingot kasvavat (Talousindikaattorit n.d.). Kuitenkin BKT:n kasvaessa liian nopeasti, inflaatio kiihtyy ja syö rahan arvoa johtaen sijoitusten arvon heikentymiseen (Talousindikaattorit n.d.). Mikrotasolla vaikuttaviin tekijöihin kuuluvat yritykset, markkinat ja sijoittajan psykologia, kuten sijoittajan mieliala sekä huolestuneisuus (Liu ja muut 2022 s. 2).

Monet sijoittajat seuraavat osakemarkkinoita hyvin tarkkaan. Sijoituspäätösten perustaminen osakkeen hinnan muutoksen voi merkittävästi pienentää riskiä. Kuitenkin osakemarkkinoiden ennustamista pidetään yhtenä vaikeimmista ongelmista aikasarjaennustamisessa. Osakemarkkinoiden ennustamisesta on paljon tutkimuksia ja niissä käytetyt metodit voidaan jakaa karkeasti kolmeen kategoriaan: perinteiset aikasarjamallit, koneoppimisalgoritmit ja neuroverkkomallit.

Perinteiset aikasarjamallit kehitettiin estimoimaan osakehintadataa käyttämällä mennyttä ja nykyistä dataa (Pokhrel ja muut 2022 s. 2). Perinteisiä aikasarjamalleja ovat esimerkiksi AR- (Autoregressive), MA- (Moving Average), ARMA- (Autoregressive and Moving Average) ja ARIMA-mallit (Autoregressive Integral Moving Average) (Liu ja

muut 2022, s. 2). Nämä tilastolliset mallit kuvaavat lineaarisia suhteita hyvin, mutta eivät ole kovin käytännöllisiä osakemarkkinoiden ennustamisessa johtuen osakedatan epälineaarisesta ja kohinaisesta käyttäytymisestä (Pokhrel ja muut 2022, s. 2).

Vastauksena tilastollisten mallien heikkouksiin on tekoäly. Koneoppimismenetelmät kykenevät perinteisistä tilastollisista malleista poiketen kuvaamaan epälineaarisia muutoksia. Nopea kehitys tekoäly- ja koneoppimistekniikoissa, laajamittaisen datan saatavuus ja kasvanut laskentakapasiteetti mahdollistavat vakaiden koneoppimismallien kehittämisen. Koneoppimismetodeja käytetään usein osakemarkkinan analyysissä ja osaketrendien ennustamisessa. Ennen nopeaa syväoppimismallien kehitystä osakkeita on ennustettu esimerkiksi tukivektorikoneella (eng. support-vector machine, SVM) (Liu ja muut 2022, s. 2).

Syväoppimismallien käyttöä on tutkittu paljon osakemarkkinoiden ennustamisessa. Useimmat artikkelit käyttävät jonkinlaista neuroverkkoarkkitehtuuria. Eniten käytettyjä syväoppimisarkkitehtuureja on LSTM, CNN, GRU ja niiden erilaiset yhdistelmät (Pokhrel ja muut 2022, s. 2). LSTM:ään ja GRU:hun tutustutaan tarkemmin luvussa 3. Viime vuosina LSTM ja GRU ovat saaneet merkittävää huomiota johtuen niiden nopeasta kehityksestä tekoälyssä (Liu ja muut 2022, s. 3). LSTM:stä ja GRU:sta kehitetyt BiLSTM ja BiGRU ovat näiden kaksisuuntaiset versiot. Niissä on yhdistetty kaksi neuroverkkoa, joista toisen tieto kulkee vastakkaiseen suuntaan. Liu ja muiden (2022) mukaan BiLSTM soveltuu epälineaarisen datan käsittelyyn paremmin kuin tavallinen LSTM. Patra ja muut (2022) tekivät mallin, jossa oli sekä GRU- että LSTM-kerros. Malli suoriutui useimmilla suoritusmittareilla paremmin kuin GRU ja LSTM.

2.2 Keskipoikkeama (MAE) ja prosentuaalisen virheen keskiarvo (MAPE)

Keskipoikkeama (Mean Absolute Error, MAE) mittaa oikean ja ennustetun datan välisen eron itseisarvon keskiarvoa eli se on kaikkien virheiden keskiarvo.

Keskipoikkeama on esitetty yhtälössä (1).

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

jossa \hat{y}_i on ennustettu datapiste ja y_i on ennustettua dataa vastaava todellinen datapiste (Patra & Mohanty 2022, s. 7).

Prosentuaalisen virheen keskiarvo (Mean Absolute Percentage Error, MAPE) eroaa keskipoikkeamasta siinä, että siinä mitataan prosentuaalisen virheen itseisarvon keskiarvoa. Virheen suuruus siis suhteutetaan todelliseen arvoon, kuten yhtälössä (2).

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

Prosentuaalisen virheen keskiarvon käytössä tulee ongelmia, jos todellinen arvo $y_i = 0$, koska silloin joudutaan jakamaan nolalla ja tulos on määrittelemätön. Lisäksi sitä ei voida käyttää asteikoilla, joissa ei ole suhteellisuutta, kuten Celsius- ja Fahrenheit-asteikoilla.

2.3 Keskineliövirhe (MSE) ja keskineliövirheen neliöjuuri (RMSE)

Keskineliövirhe (Mean Squared Error, MSE) mittaa virheen neliön keskiarvoa eli se mittaa virheen suuruutta. Se saadaan kun lasketaan todellisen datan ja ennustetun datan erotuksen neliö jokaisesta datapisteestä ja lasketaan näiden datapisteiden keskiarvo, kuten yhtälössä (3).

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3)$$

Keskineliövirhe on kuin keskipoikkeama, mutta siinä otetaan virheen sijaan virheen neliö. Keskineliövirhe siis antaa suuremman painon niille arvoille, jotka poikkeavat paljon todellisesta arvosta. Joissain tilanteissa se voi vääristää tulosta.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

Keskineliövirheen neliöjuuri (Root Mean Squared Error, RMSE) saadaan ottamalla keskineliövirheestä neliöjuuri, kuten yhtälössä (4). RMSE kuvaa MSE:n tavoin virheen suuruutta. RMSE:n etu on siinä, että siitä saatavat lukuarvot ovat samassa suuruusluokassa kuin MAE:ssa, joten sitä on helpompi tulkita kuin MSE:tä.

2.4 Selitysaste

Mallin selitysaste eli R^2 kuvaa mallin yhteensopivuutta otoksen kanssa (Shahi ja muut 2020, s. 7). Se annetaan usein prosenttilukuna. Yhtälössä (5) on esitetty miten selitysaste lasketaan.

$$R^2 = 1 - \frac{\sum_{d=1}^N (a_d - p_d)^2}{\sum_{d=1}^N (a_d - \bar{a})^2} \quad (5)$$

$$\bar{a} = \frac{1}{N} \sum_{d=1}^N a_d \quad (6)$$

jossa p_d on ennustettu hinta ja a_d on todellinen hinta aikasarjan päivänä $d=1..N$.

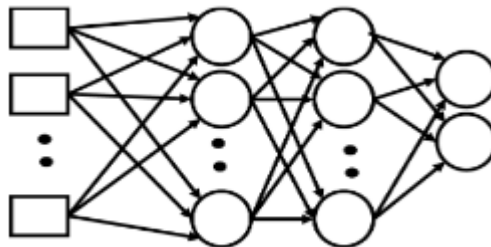
Yhtälöstä (6) voidaan havaita, että yhtälössä (5) esiintyvä \bar{a} on todellisten hintojen keskiarvo.

3. TAKAISINKYTKETYT NEUROVERKOT

Tässä luvussa kerrotaan syväoppimisesta ja määritellään tutkielmassa käsiteltäviin syväneuroverkkoihin liittyvät keskeisimmät käsitteet, kuten RNN, LSTM ja GRU.

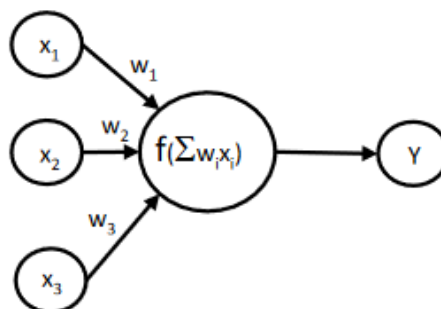
3.1 Neuroverkot ja syväoppiminen

Neuroverkko on oppimispohjainen informaation käsittelyn malli, johon on saatu idea ihmisaivojen hermoverkoista. Neuroverkko on nimensä mukaisesti neuroneista koostuva verkko, jonka rakenne on esitetty kuvassa 1. Siinä neliömuotoiset neuronit esittävät syötekerrosta, mihin piirteet syötetään ja pyöreänmuotoiset solmut ovat kaksi piilokerrosta ja ulostulokerros. Kahden neuronin välinen kytkentä on nimeltään synapsi.



Kuva 1: Täysin kytketty neuroverkko (Pietikäinen & Silven 2019, s. 85)

Kuvassa 2 on kuvattu yksittäisen neuronin rakenne. Siinä neuroniiin tulevat syötteet x_1 , x_2 ja x_3 , joilla jokaisella on synapsin mukainen painokerroin (w_1 , w_2 ja w_3). Näistä lasketaan neuronissa painotettu summa. Painotetusta summasta lasketaan aktivointifunktiolla, kuten sigmoid- tai askelfunktiolla, epälineaarinen kuvaus (Pietikäinen & Silven 2019, s. 87). Aktivointifunktio antaa ulostulona joko arvon 0 tai 1, joka kuvaa neuronin aktivoitumista. Neuroverkkojen koulutus tarkoittaa käytännössä oikeiden painokertoimien löytämistä synapseille.

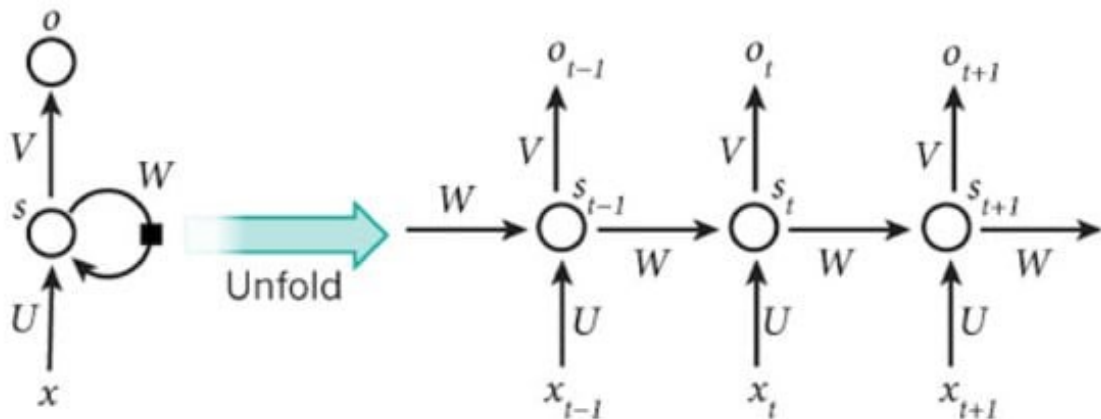


Kuva 2: Yksittäinen neuroni (Pietikäinen & Silven 2019, s. 85)

Yleensä painokertoimet määritellään vastavirta-algoritmeilla. Algoritmin nimi tulee siitä, että siinä aloitetaan neuroverkon ulostulokerroksesta ja edetään kohti syötekerrosta. Ensin selvitetään ulostulopään ensimmäisten synapsien painokertoimien virheet ja katotaan mikä vaikutus edellisen kerroksen neuroneilla oli virheeseen ja näin edetään kerros kerrallaan. Lopulta tiedetään, miten kukin neuroni vaikuttaa kokonaisvirheeseen ja muutetaan kutakin painokerrointa siten, että kokonaisvirhe pienenee (Pietikäinen & Silven 2019, s. 85-86). Kouluttautumisvaiheessa lasketaan virheen suuruus häviöfunktion avulla. Häviöfunktiona voidaan käyttää esimerkiksi luvussa 2.2 esiteltyä keskineliövirhettä.

3.2 Takaisinkytketyt neuroverkot (RNN)

Tavallinen syväneuroverkko ei voi oppia yksittäisten datanäytteiden keskinäisiä riippuvuuksia (Pietikäinen & Silven 2019, s. 98). Tämä on ongelma kaikessa sekventiaalisessa datassa, esimerkiksi kielen käsittelyyn perustuvissa sovelluksissa ja osakemarkkinadatan analysoimisessa. Tällaisissa sovelluksissa voidaan hyödyntää takaisinkytkettyjä neuroverkkoja (RNN). RNN on neuroverkkoarkkitehtuuri, joka kykenee käsittelemään aikasarjadataa, kuten osakemarkkinadataa (What are recurrent networks? 2020). Takaisinkytkettyjen neuroverkkojen tutkiminen alkoi 1980-luvulla, ja 2000-luvun aikana niistä kehittyi yksi keskeisistä syväoppimisalgoritmeista (Zhou ja muut 2022).



Kuva 3: Yksinkertainen RNN-arkkitehtuuri (Dey ja muut 2021 s. 3)

Kuvassa 3 on kaksi eri tapaa visualisoida RNN. Vasemmalla oleva visualisointi on tiiviimpi esitysmuoto RNN:stä. Siinä RNN:n syöte on x , joka kuvaa koko syötettävää sarjaa ja ulostulo on o , joka kuvaa koko ulostulosarjaa. Oikealla oleva visualisointi kuvaa yksittäisiä aika-askelia. Siinä syötteet x_{t-1} , x_t ja x_{t+1} kuvaavat peräkkäisiä syötteitä ja o_{t-1} , o_t ja o_{t+1} peräkkäisiä ulostuloja. W , U ja V kuvaavat neuroverkon painoja.

Jokaisessa aika-askeleessa on piilokerros s , jolle lasketaan ulostulo halutulla aktivointifunktiolla, esimerkiksi hyperbolisella tangentilla, kuten yhtälössä (7)

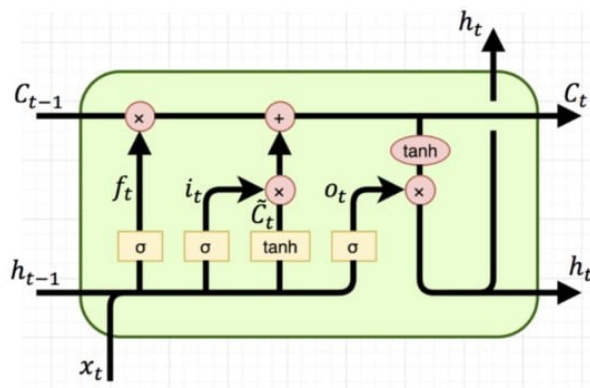
$$S_t = \tanh(Wx_t + US_{t-1} + b) \quad (7)$$

missä b on vakiotermi.

Tavallinen RNN kärsii häviävän gradientin ja hajaantuvan gradientin ongelmista, minkä vuoksi RNN:n on vaikea oppia syy-seuraussuhteita (Banerjee 2022, s. 29). Häviävän gradientin ongelmassa painoja päivittävät parametrit alkavat lähestyä nollaa, jolloin neuroverkon oppiminen eli painojen päivittyminen hidastuu tai loppuu kokonaan (What are recurrent networks? 2020). Hajaantuvan gradientin ongelmassa taas painot lähtevät kasvamaan liian suuriksi luoden epävakaan mallin (What are recurrent networks? 2020).

3.3 LSTM-neuroverkko

LSTM on RNN-pohjainen malli, jota käytetään kun pitkän ajan riippuvuudet ovat merkittävä osa oppimisprosessia (Alkhatib ja muut 2022 s. 7). LSTM siis on luotu ratkomaan hajaantuvan gradientin ongelma (Dey ja muut 2021 s. 4).



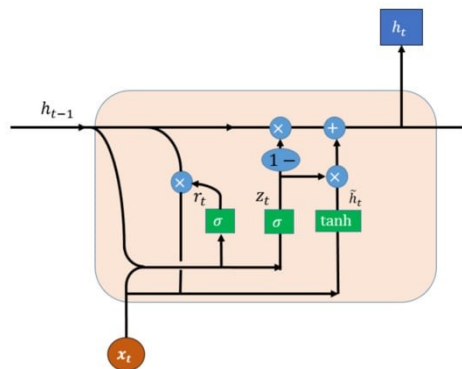
Kuva 4: Yksisoluinen LSTM-arkkitehtuuri (Dey ja muut 2021 s. 4)

Kuvassa 4 on esitetty yksittäinen LSTM-neuroni. LSTM koostuu viidestä osasta; solutilasta (C_t), unohdusportista (f_t), syöteportista (i_t), ulostuloportista (o_t) ja piilotetusta tilasta (h_t). Solutila C_{t-1} on yksiulotteinen vektori, joka sisältää muistissa olevan tiedon. Solutila kulkee LSTM-solun läpi, jossa unohdusportti ja syöteportti muokkaavat sitä. Unohdusportti f_t pyrkii eliminoimaan solutilasta aikaisempien aika-askelten arvot, joita ei tarvita. Syöteportti i_t pyrkii lisäämään solutilaan uutta tietoa nykyisestä aika-askeleesta. Näiden muokkausten jälkeen saadaan uusi solutila C_t . Lopuksi päätetään

uusi piilotettu tila h_t , joka riippuu uudesta solutilasta C_t ja vanhasta piilotetusta tilasta h_{t-1} . Käytännössä tässä päätetään, mitä osia solutilasta halutaan antaa ulostulona.

3.4 GRU-neuroverkko

GRU on muunnelma RNN:stä, joka LSTM:n tavoin ratkaisee RNN:n ongelmat. Se eroaa kuitenkin LSTM:stä siten, että se yhdistää unohdusportin ja syöteportin yhdeksi portiksi nimeltä päivitysportti (Alkhatib ja muut 2022, s. 8). Päivitysportti hoitaa sekä unohdusportin että syöteportin tehtävät, eli se päättää mitä tietoa lisätään ja mitä poistetaan. GRU-solun rakenne on esitetty kuvassa 5. Se on rakenteeltaan yksinkertaisempi, joten se on nopeampi kouluttaa.



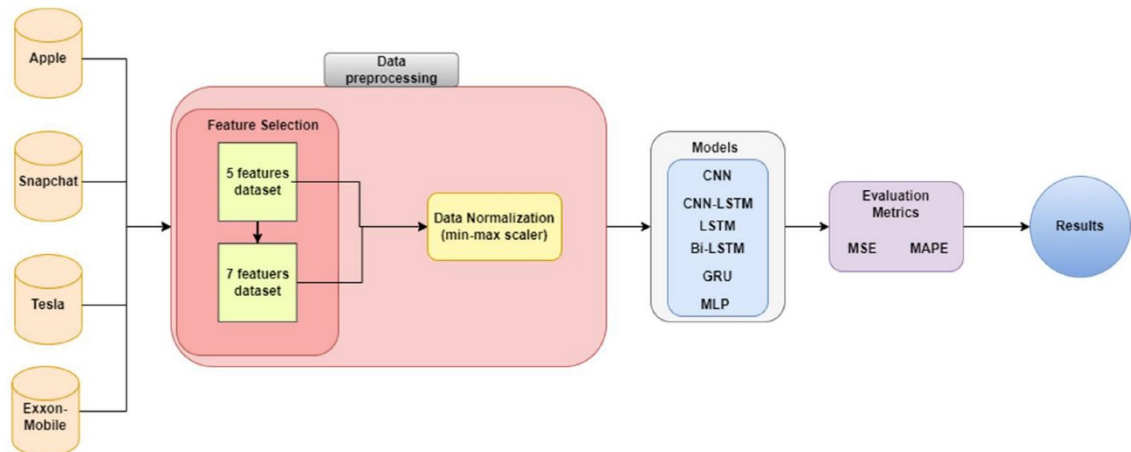
Kuva 5: Yksittäisen GRU-solun arkkitehtuuri (Dey ja muut 2021, s. 5)

4. SYVÄNEUROVERKOT OSAKEKURSSIEN ENNUSTAMISESSA

Tässä luvussa on esitellään miten LSTM:ää ja GRU:ta käytetään osakekurssien ennustamiseen eri tutkimuksissa ja analysoida mallien luotettavuutta.

4.1 Käytetyt syväoppimismallit

Alkhatib ja muut (2022) vertasivat tutkimuksessaan muutamaa syväoppimismallia osakehintojen ennustamiseen, joista tässä tutkielmassa tarkastellaan vain GRU:ta ja LSTM:ää. Tutkimuksessa haluttiin verrata miten neuroverkoille syötettävien piirteiden määrä ja datajoukon suuruus vaikuttavat ennustustarkkuuteen.



Kuva 6: vuokaavio osakkeiden ennustamisprosessista (Alkhatib ja muut, s. 6)

Kuvassa 6 on havainnoillistettu tutkimusprosessia. Dataa on käytetty neljästä yrityksestä, Applelta, Snapchatilta, Teslalta ja ExxonMobilelta. Applelta ja ExxonMobilelta dataa on 21 vuoden ajalta, Teslalta 11 vuodelta ja Snapchatilta 4 vuodelta. Neuroverkot on koulutettu ensin käyttäen neljää syötepiirrettä:

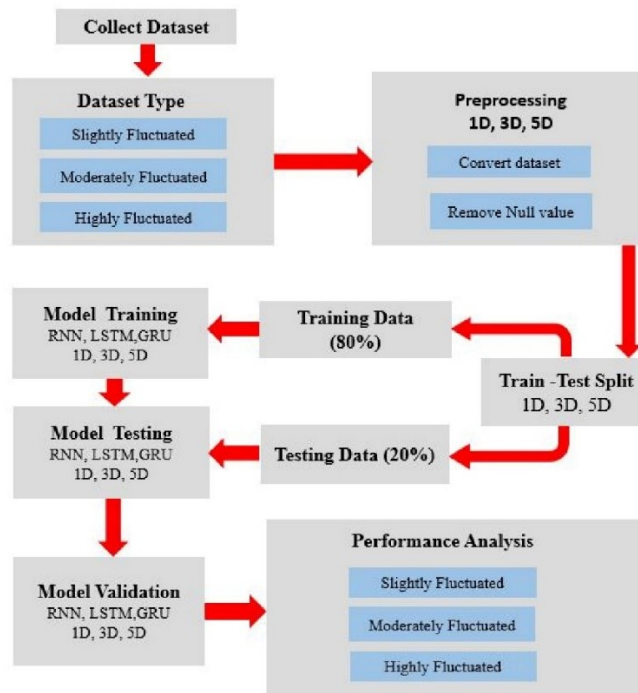
- High: osakkeen korkein hinta kyseisenä päivänä
- Low: osakkeen alhaisin hinta kyseisenä päivänä
- Open: osakkeen avaushinta kyseisenä päivänä
- Volume: kuinka paljon on käyty kauppaa kyseisenä päivänä

Näiden neljän syötepiirteiden pohjalta on luotu 2 syötepiirrettä lisää, High-Low ja Open-Close. High-Low tarkoittaa päivän korkeimman ja alhaisimman hinnan erotusta ja Open-Close päivän aloitus- ja päätöskurssien erotusta. Tämän jälkeen data on normalisoitu

välille $[0,1]$. Datan esikäsittelyn jälkeen data on syötetty neuroverkoille käyttäen häviöfunktiona MSE:tä ja MAPE:a.

Tutkimuksessa käytetyllä LSTM:llä oli kolme kerrosta, syötekerros, 32 LSTM-neuronin piilokerros ja ulostulokerros. Käytetyllä GRU:lla oli neljä kerrosta, syötekerros, kaksi piilokerrosta, joista ensimmäisessä oli 50 ja toisessa 25 GRU-neuronia sekä ulostulokerros. Molemmissa neuroverkoissa oli aktivointifunktiona ELU ja optimointifunktiona Adam. LSTM:ää koulutettiin 100 jakson verran eräkoon 2 ja GRU:ta koulutettiin 70 jakson verran eräkoon ollessa 2.

Dey ja muut (2021) vertasivat tutkimuksessaan kolmea takaisinkytkettyä neuroverkkoa sijoitusten ennustamisessa, LSTM:n ja GRU:n lisäksi yksinkertaista RNN:ää. Lisäksi tutkimuksessa on selvitetty miten tuloksiin vaikuttaa 1, 3 ja 5 päivän aikaikkunan käyttäminen. Tutkimuksessa käytettiin kolmea datajoukkoa, jotka eroavat vaihtelultaan. Paljon vaihtelevalla osakkeella tarkoitetaan tässä osaketta, jonka hinnassa on paljon odottamatonta nousua ja laskua. Hieman vaihtelevana datajoukkona on käytetty Honda Motor Companyn (HMC) dataa, maltillisesti vaihtelevana datajoukkona Oracle Corporationin (ORCL) dataa ja suuresti vaihtelevana datajoukkona Intuit Inc.:in (INTU) dataa. Kaikista datajoukoista kerättiin dataa 20 vuodelta (30.6.2000-21.7.2020) ja jokaisessa datajoukossa on 5044 ilmentymää eli 5044 päivän tiedot. Kaikissa kolmessa datajoukossa on seitsemän attribuuttia: päivämäärä, aloituskurssi, korkein hinta, alhaisin hinta, päätöskurssi ja vaihdon volyyymi.



Kuva 7: Järjestelmäarkkitehtuuri osakehintojen ennustamiseksi eri aikaikkunoille (Dey ja muut 2021, s. 6)

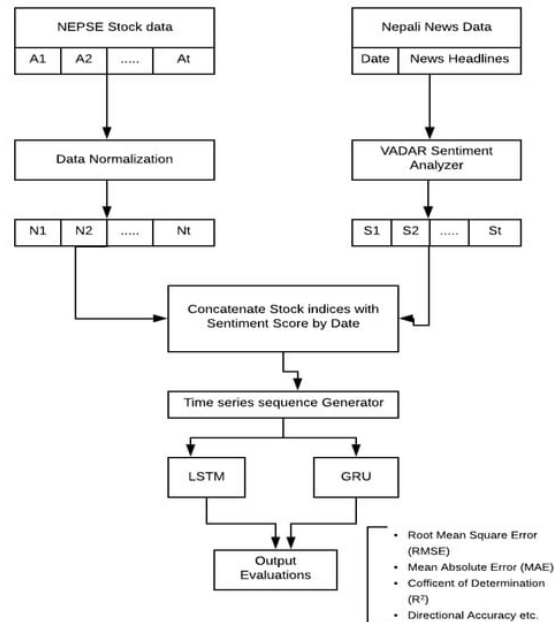
Kuvassa 7 on esitelty miten Dey ja muut (2021) ovat käsitelleet dataa. Datajoukot on luokiteltu sen mukaan, kuinka paljon ne vaihtelevat. Datan esikäsittelynä on yhden päivän aikavälin mallissa täydennetty tyhjät arvot laskemalla keskiarvo 30 edelliseltä päivältä. 3 ja 5 päivän aikaväleille on laskettu omat datajoukkonsa seuraavalla tavalla:

- Päivämäärä: ensimmäinen päivä 3 tai 5 päivän aikaikkunasta
- Aloituskurssi: 3 tai 5 päivän aikaikkunan ensimmäisen päivän aloituskurssi
- Korkein hinta: koko 3 tai 5 päivän aikaikkunan korkein hinta
- Alhaisin hinta: koko 3 tai 5 päivän aikaikkunan matalin hinta
- Päätöskurssi: 3 tai 5 päivän aikaikkunan viimeisen päivän päätöskurssi
- Vaihdon volyyymi: koko 3 tai 5 päivän aikaikkunan vaihdon volyyymi

Datan esikäsittelyn jälkeen se on jaettu koulutus- ja testausdataan. Neuroverkot optimoitiin Adamilla ja häviöfunktiona käytettiin MSE:tä. Piilokerroksia on kaksi joka mallissa, ja niiden neuronien määrä riippuu käytetystä datajoukosta ja aikaikkunasta.

Shahi ja muut (2020) pyrkivät tutkimuksessaan vertaamaan LSTM:n ja GRU:n ennustustarkkuutta mahdollisimman samoissa olosuhteissa ja objektiivisesti. Tutkimus on rajattu koskemaan nepalilaisia uutisia ja osakemarkkinadataa. Lisäksi tutkimuksessa haluttiin verrata, miten ennustustarkkuuteen vaikuttaa, kun perinteisen

osakemarkkinadatan lisäksi ennusteissa hyödynnetään finanssialan uutisten tunteita syötteenä.

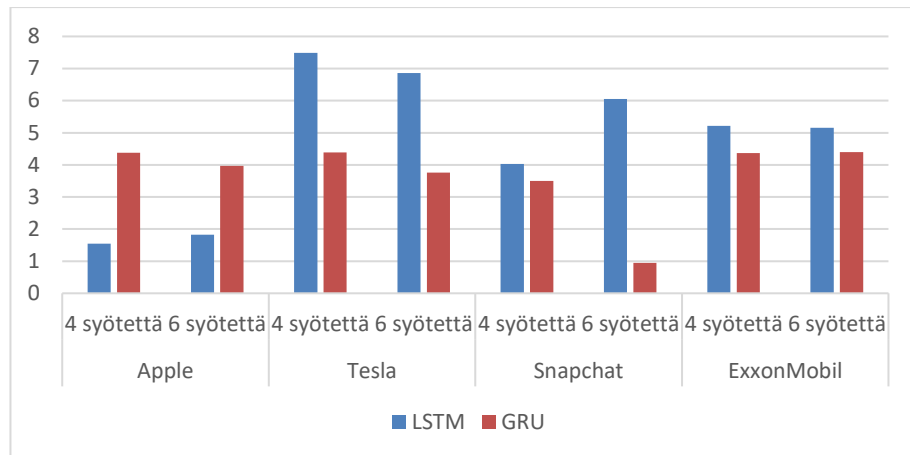


Kuva 8: järjestelmäarkkitehtuuri osakemarkkinoiden ennustamisesta, jossa yhdistyy osake- ja uutisdata (Shahi ja muut 2020, s. 4)

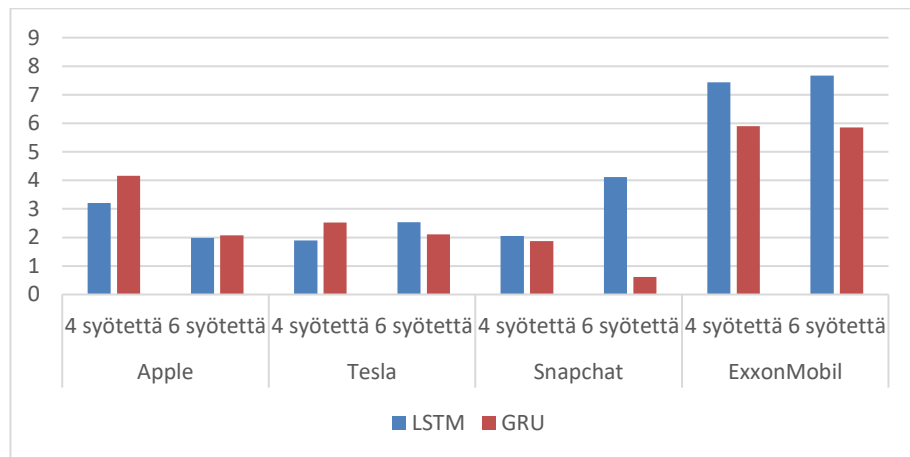
Molemmat LSTM ja GRU on koulutettu käyttämällä MAE:a häviöfunktiona. Osakedatajoukkona on käytetty ADBL:n (agriculture development bank) dataa aikavälillä 20.3.2011-14.11.2019. Osakedatajoukon attribuutit ovat aloituskurssi, päätöskurssi, korkein hinta, alhaisin hinta ja volyyymi. Samalta aikaväliltä on käytetty 42 110 uutista, joista on otettu attribuuteiksi päivämäärä, uutisen otsikko ja uutisen leipäteksti.

Jotta LSTM:n ja GRU:n vertailu olisi mahdollisimman objektiivista, tutkimuksessa on käytetty molemmilla samaa määrää kerroksia, samaa aktivointifunktiota, hyperbolista tangenttia, ja samoja syötteitä. Joka mallissa on ensin yksi syötekerros, yksi LSTM- tai GRU-kerros, yksi pudotuskerros sekä lopuksi ulostulokerros. Molempia koulutetaan enintään 100 epokin verran aikaisen pysähtymisen (engl. early stopping) metodia hyödyntäen eräkokona 30 näytettä. Arviointimittareina on käytetty mm. MAE:ta ja RMSE:tä.

4.2 Mallien tarkkuus ennustamisessa

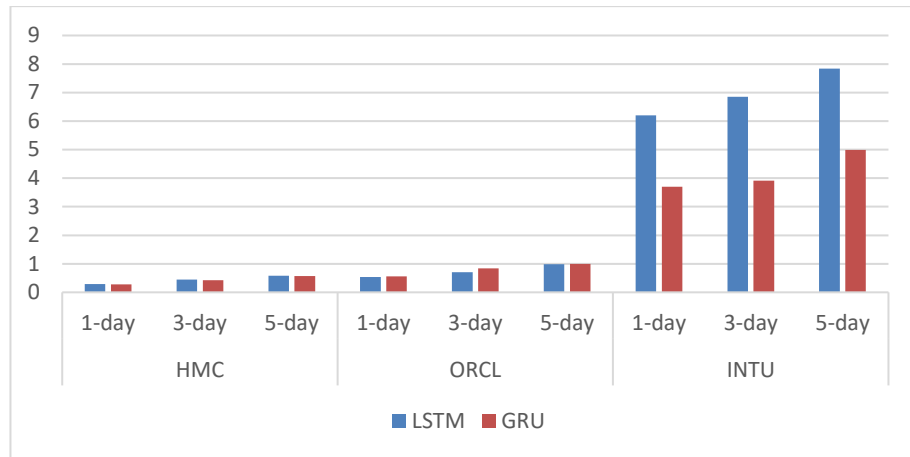


Kuva 9: keskineliövirheen neliöjuuri (RMSE) eri datajoukoilla ja syötepiirteillä (Alkhatib ja muut 2022)

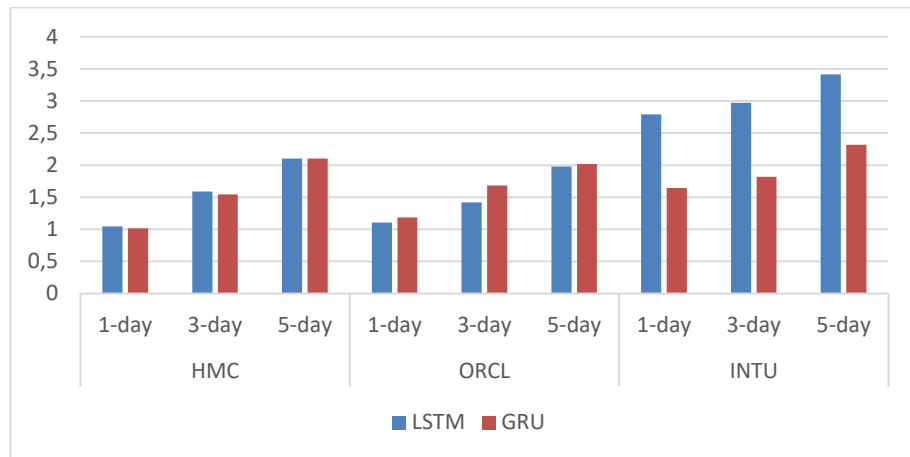


Kuva 10: Prosentuaalisen virheen keskiarvo (MAPE) eri datajoukoilla ja syötepiirteillä (Alkhatib ja muut 2022)

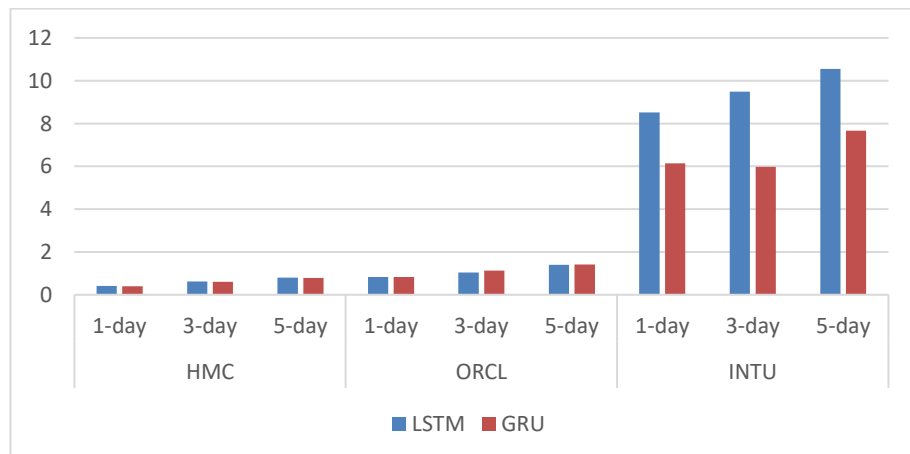
Kuvissa 9 ja 10 on esitelty Alkhatib ja muiden tulokset. RMSE:llä mitattuna GRU voittaa LSTM:n kaikilla muilla datajoukoilla paitsi Applella, kun taas MAPE:lla mitattuna GRU voittaa ainoastaan osassa tapauksista. Syötepiirteiden määrän lisääminen näyttäisi parantavan enemmän GRU:n kuin LSTM:n tuloksia. Kuvan 9 perusteella voisi tulkita, että GRU ja LSTM ennustavat Teslan dataa huonoiten, kun taas kuvan 10 perusteella ExxonMobilia. Ero johtuu kuitenkin siitä, että RMSE ja MAPE painottavat virheitä eri tavalla. RMSE ottaa virheestä neliön, joten se korostaa suurempia virheitä enemmän. RMSE:llä mitattuna GRU pärjää paremmin kuin LSTM, kun taas MAPE:lla mitattuna kumpikaan ei ole selkeästi ylivoimainen. GRU:n tuloksissa on siis todennäköisesti vähemmän suuria yksittäisiä virheitä kuin LSTM:llä. Pienemmillä datajoukoilla, Teslalla ja Snapchatilla, GRU vaikuttaisi pärjäävän paremmin. Sen sijaan suurempien datajoukkojen tapauksessa, Applella ja ExxonMobililla, ei ole selkeää voittajaa.



Kuva 11: Keskipoikkeama (MAE) eri datajoukoilla ja aikaikkunoilla (Dey ja muut 2021)



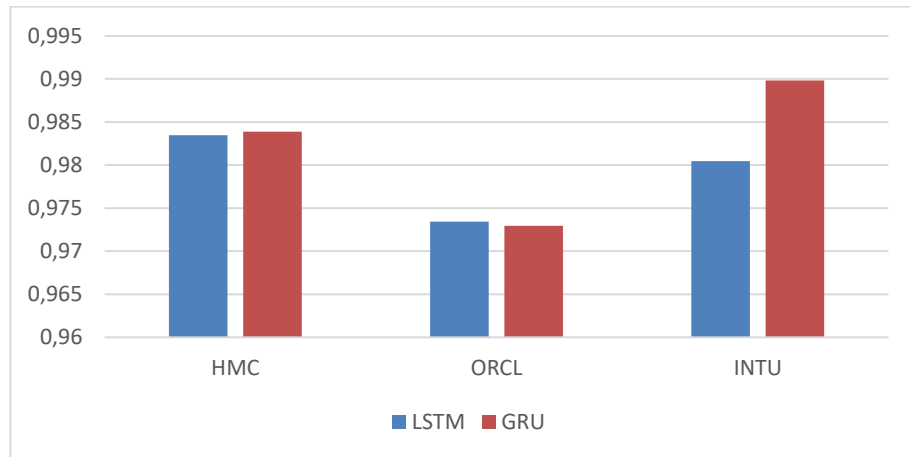
Kuva 12: Prosentuaalinen virheen keskiarvo (MAPE) mitattuna eri datajoukoilla ja aikaikkunoilla (Dey ja muut 2021)



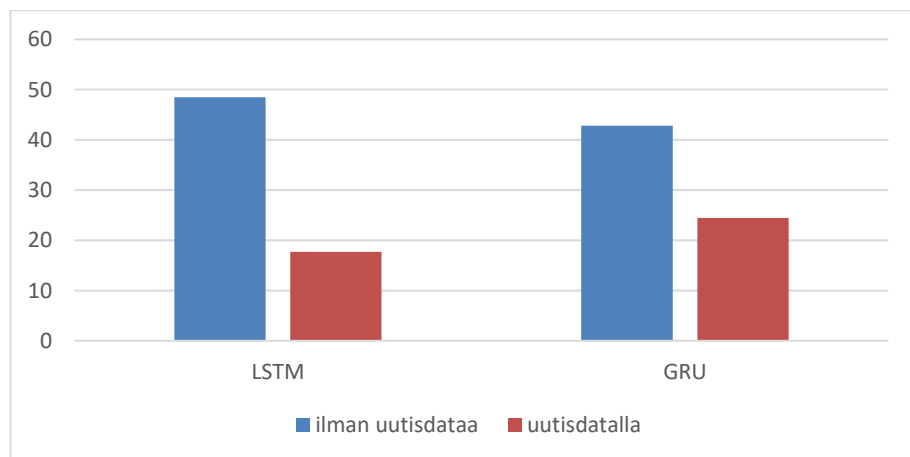
Kuva 13: Keskineliövirheen neliöjuuri (RMSE) mitattuna eri datajoukoilla ja aikaikkunoilla (Dey ja muut 2021)

Kuvissa 11-14 on esitetty Dey ja muut (2021) tulokset. Kuvista voidaan havaita, että lyhyemmän aikavälin käyttö antaa paljon tarkempia tuloksia, eikä 3 ja 5 päivän

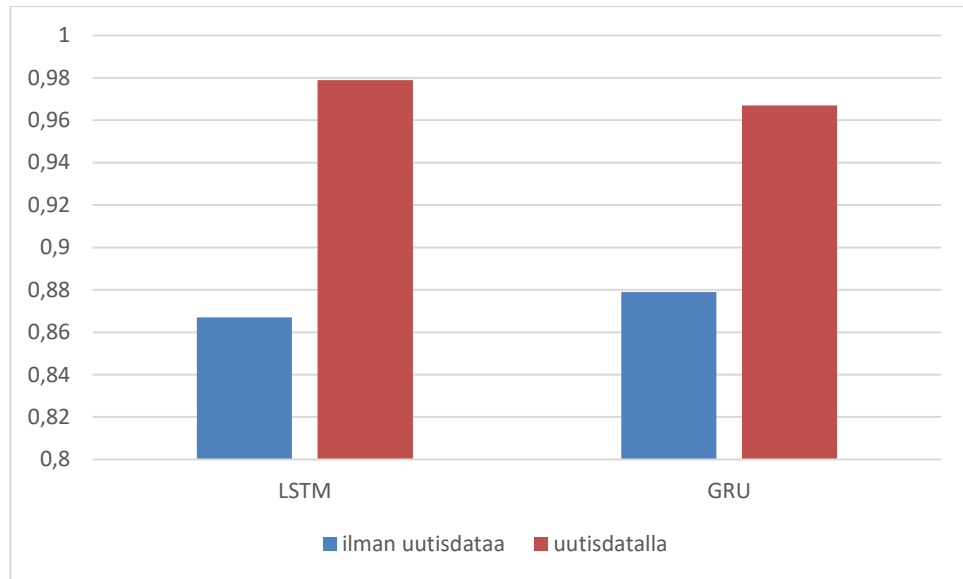
aikaikkunoista ole hyötyä. Tässä tutkimuksessa kaikilla mittareilla mitattuna GRU saavuttaa LSTM:ää parempia tuloksia vaihtelultaan pienimmällä Honda Motors Companylla ja vaihtelultaan suurimmalla Intuit Inc:llä, LSTM:n saavuttaessa GRU:ta parempia arvoja maltillisesti vaihtelevalla Oracle Corporationilla.



Kuva 14: Dey ja muut (2021) yhden päivän aikavälin selitysaste eri datajoukoilla



Kuva 15: Keskipoikkeama (MAE) mitattuna uutisdatalla ja ilman (Shahi ja muut 2020)



Kuva 16: selitysaste (R^2) mitattuna ilman uutisdataa ja uutisdatalla (Shahi ja muut 2020)

Kuvissa 14 ja 15 on esitetty Shahi ja muut (2020) tulosten keskiarvo ja selitysaste. Kuvasta 5 voidaan tulkita, että kun neuroverkolle syötetään historiallisen osakedatan lisäksi uutisista saatua sentimenttidataa, virheet vähenevät merkittävästi. GRU suoriutuu pelkällä historiallisella osakedatalla paremmin kuin LSTM, mutta kun otetaan uutisista analysoitu sentimenttidata mukaan koulutukseen, LSTM voittaa GRU:n.

Kaikista tutkimuksista voidaan vetää yhteinen johtopäätös, ettei kumpikaan syväoppimismalli, LSTM tai GRU, ole ylivoimainen osakehintojen ennustamisessa.

Lisäksi täysin objektiivinen neuroverkkomallien vertailu on vaikeaa. Esimerkiksi Alkhatib ja muut (2022) ja Dey ja muut (2021) ovat käyttäneet eri kokoisia neuroverkkoja. Alkhatib ja muut (2022) ovat käyttäneet syvempää GRU:ta, mutta neuroverkkojen kokoerolle ei ole annettu syytä. Dey ja muut (2021) ovat valinneet neuroverkon koon käytettävän aikaikkunan sekä syväoppimismallin perusteella. Kaikissa heidän kouluttamissaan verkoissaan on kaksi piilokerrosta, mutta eri määrä neuroneja. Shahi ja muut (2020) ovat halunneet tehdä vertailusta mahdollisimman objektiivisen käyttämällä saman kokoisia verkkoja. Kuitenkaan kaikille neuroverkoille sama koko ei välttämättä ole se optimaalisin.

Ennustustarkkuuteen vaikuttaa käytettävä datajoukko. Lisäksi se, miten neuroverkko on koulutettu, on tärkeää. Syötettävien attribuuttien määrä ei automaattisesti tarkenna tuloksia. Alkhatib ja muut (2022) tutkimuksessa kaksi uutta syötepiirrettä ei tuonut merkittävää lisäinformaatiota neuroverkoille. Sen sijaan aikaikkunan muokkaaminen jätti paljon oleellista informaatiota ja heikensi ennustustarkkuutta. Uutisdatan

hyödyntäminen historiallisen datan apuna paransi tuloksia merkittävästi. Kuitenkin verrattaessa kuvia 14 ja 16, voidaan havaita, että Dey ja muut (2020) on päässyt melko samoihin tuloksiin ilman uutisdataa kuin Shahi ja muut (2020) käyttämällä uutisdataa. Tästä voidaan päätellä, että uutisdataa hyödyntävä malli ei automaattisesti saavuta parempia tuloksia kuin malli, joka perustuu ainoastaan historiallisen osakedatan käyttöön.

YHTEENVETO

Tämän tutkielman tarkoituksena oli selvittää, miten takaisinkytkettyjä neuroverkkoja, LSTM:ää ja GRU:ta koulutetaan ennustamaan osakedataa ja eroavatko näiden neuroverkkojen ennustusten luotettavuus toisistaan merkittävästi sekä mitkä tekijät vaikuttavat ennustusten luotettavuuteen.

Takaisinkytkettyjä neuroverkkoja koulutetaan ennustamaan syöttämällä osakkeista aikasarjadataa pitkältä aikaväliltä, yleensä 20 vuoden dataa. Tyypillisesti neuroverkolle annetaan syötteenä jonkin osakkeen päivän avaushinta, korkein arvo, matalin arvo ja myynnin volyyymi jokaiselta valitun aikavälin päivältä, mutta syötepiirteet voivat vaihdella. Lisäksi uutisdataa hyödynnettäessä lasketaan uutisten otsikoiden ja leipätekstin perusteella uutiselle numeerinen arvo, joka kertoo onko uutinen positiivinen, negatiivinen vai neutraali.

LSTM:n ja GRU:n vertailussa kävi ilmi, ettei kumpikaan ole osakekurssien ennustustehtävässä ylivoimainen ja että niiden välillä on vaikea tehdä objektiivista vertailua. Ennustusten tarkkuuteen vaikuttaa eniten tehdyt valinnat neuroverkkojen kouluttamisessa. Dataa tulee olla riittävältä aikaväliltä, mutta data-attribuuttien lisääminen ei automaattisesti tuo neuroverkolle mitään oleellista informaatiota, jos ne on valittu huonosti. Oleellisen datan jättäminen voi toisaalta heikentää tuloksen tarkkuutta. Uutisdata taas näyttää parantavan tuloksia merkittävästi.

Jatkotutkimuskohteina voitaisiin selvittää, onko takaisinkytkettyillä neuroverkoilla enää käyttöä tulevaisuudessa. Osakekurssien kehityksen ennustamisessa etsitään jatkuvasti parempia malleja. Esimerkiksi 2017 esitellyt transformerit on suunniteltu takaisinkytkettyjen neuroverkkojen tavoin käsittelemään aikasarjadataa, mutta pystyvät käsittelemään kokonaisen syötteen kerralla, mikä pienentää neuroverkon koulutusaikaa. Tämän tutkielman ulkopuolelle jäivät myös kokonaan esimerkiksi konvoluutioneuroverkot sekä erilaiset variaatiot LSTM:stä ja GRU:sta, kuten niiden kaksisuuntaiset versiot.

LÄHTEET

- Alkhatib, K., Khazaleh, H., Alkhazaleh H. A., Alsoud, A. R. & Abualigah L. (2022). A New Stock Price Forecasting Method Using Active Deep Learning Approach. *Journal of Open Innovation*, Vol.8(2), 23 p.
- Banerjee, S. & Mukherjee, D. (2022). Short Term Stock Price Prediction in Indian Market: A Neural Network Perspective. *Studies in Microeconomics*, 10(1), 27 p.
- Dey, P., Hossain, E. Hossain, M. I., Chowdhury, M. A., Alam, M. S., Hossain, M. S. & Andersson, K. (2021). Comparative analysis of recurrent neural networks in stock price prediction for different frequency domains. *Algorithms*, 14(8), 20 p.
- IBM Cloud Education (2020). What are recurrent neural networks? Saatavissa (viitattu 1.11.2022): <https://www.ibm.com/cloud/learn/recurrent-neural-networks>
- Lawi, A., Mesra, H. & Amir, S. (2022). Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately. *Journal of Big Data*, 9(1), 1–19.
- Li, M. Zhu, Y., Shen, Y., & Angelova, M. (2022). Clustering-enhanced stock price prediction using deep learning. *World Wide Web (Bussum)*, p. 1–26.
- Liu, B., Yu, Z., Wang, Q., Du, P. & Zhang, X. (2022). Prediction of SSE Shanghai Enterprises index based on bidirectional LSTM model of air pollutants. *Expert Systems with Applications*, 204, p. 1-9
- Lv, J., Wang, C., Gao, W. & Zhao, Q. (2021). An Economic Forecasting Method Based on the LightGBM-Optimized LSTM and Time-Series Model. *Computational Intelligence and Neuroscience*, 2021, Vol.2021, p. 1–10
- Patra G. R. & Mohanty, M. N. (2022). An LSTM-GRU based hybrid framework for secured stock price prediction. *Journal of Statistics & Management Systems*, Vol. 25(6), 20 p.
- Pietikäinen M. & Silven O. (2019). Tekoälyn haasteet: koneoppimisesta ja konenäöstä tunnetekoälyyn. Saatavissa (viitattu 13.11.2022): <http://urn.fi/urn:isbn:9789526224824>
- Pokhrel, N. R., Dahal, K. R., Rimal, R., Bhandari, H. N., Khatri, R. K. C., Rimal, B. & Hahn, W. E. (2022). Predicting NEPSE index price using deep learning models. *Machine Learning with Applications*, 9, 13 p.
- Shahi, T. B., Shrestha, A., Neupane, A. & Guo, W. (2020). Stock price forecasting with deep learning: A comparative study. *Mathematics (Basel)*, Vol. 8(9), 15 p.
- Talousindikaattorit. Julkaisuaika tuntematon. Sijoittaja.fi. Saatavissa (viitattu 19.12.2022): <https://www.sijoittaja.fi/markkinaymparisto/talousindikaattorit/>
- Wang X., Yang, K. & Liu, T. (2021). Stock Price Prediction based on Morphological Similarity Clustering and Hierarchical Temporal Memory. *IEEE Access*, 9,

Zhou, Q., Zhou, C. & Wang, X. (2022). Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection. *PloS One*, 17(2), 8 p.