

Weerti Montonen

**Toistuvien neuroverkkojen vertailu osakehintojen
ennustamisessa**

Tietotekniikan kandidaatintutkielma

22. tammikuuta 2024

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Weerti Montonen

Yhteystiedot: montwepa@student.jyu.fi

Ohjaaja: Tytti Saksa

Työn nimi: Toistuvien neuroverkkojen vertailu osakehintojen ennustamisessa

Title in English: Comparison of recurrent neural networks in predicting stock prices

Työ: Kandidaatintutkielma

Sivumäärä: 19+0

Tiivistelmä: Tämä tutkielma vertailee pitkä lyhytaikainen muisti (LSTM, engl. *long short-term memory*) ja aidatut toistuvat yksiköt (GRUs, engl. *gated recurrent units*) toistuvia neuroverkkoja ja niiden suorituskykyä osakehintojen ennustamisessa. Tarkastelluissa tutkimuksissa LSTM ja GRU pärjäsivät pitkälti yhtä hyvin. LSTM osoitti GRU:ta parempaa suorituskykyä niissä tutkimuksissa, joissa käytettiin enemmän teknisiä indikaattoreita. Tarkastelluissa tutkimuksissa, jossa tutkijat esittivät oman mallinsa, kuten WLSTM+*attention* malli, pärjäksi tutkijoiden malli molempia GRU- ja LSTM mallia paremmin.

Avainsanat: LSTM, GRU, Toistuvat neuroverkot, osakehintojen ennustaminen

Abstract: This study compares the performance of long short-term memory (LSTM) and gated recurrent units (GRU) recurrent neural networks in predicting stock prices. Based on the examined papers, both models show fairly similar performance, with LSTM outperforming GRU in studies that used more technical indicators. In the papers where the researchers presented their own model, such as the WLSTM+*attention* model, the researchers' model outperformed both the GRU and LSTM models.

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

Kuviot

Kuvio 1. Yksinkertainen neuroverkko. (Fausett 1994)	3
---	---

Taulukot

Taulukko 1. Alphabetin osakeen tiedot 9.10.2009, valuutta USD (Yahoo-Finance)	8
Taulukko 2. LSTM- ja GRU-mallien vertailu DJIA-datsetissä (Qiu, Wang ja Zhou 2020) 9	
Taulukko 3. LSTM- ja GRU-mallien vertailu S&P 500 -datsetissä (Qiu, Wang ja Zhou 2020)	10
Taulukko 4. LSTM- ja GRU-mallien vertailu HSI-datsetissä (Qiu, Wang ja Zhou 2020) .	10
Taulukko 5. LSTM- ja GRU-mallien vertailu S&P 500 -datsetissä (Sethia ja Raut 2019)	10
Taulukko 6. LSTM- ja Wavelet Transform Long Short-Term Memory with Attention Mechanism-mallien vertailu S&P 500 -datsetissä (Qiu, Wang ja Zhou 2020).....	12

Sisällys

1	JOHDANTO	1
2	NEUROVERKOT	2
	2.1 Toistuvat neuroverkot	3
	2.2 Pitkä lyhytaikainen muisti (LSTM, engl. <i>long short-term memory</i>)	4
	2.3 Aidatut toistuvat yksiköt (GRUs, engl. <i>gated recurrent units</i>)	4
3	OSAKEMARKKINAT	6
	3.1 Yritys ja osakkeet	6
	3.2 Osakkeiden ennustaminen neuroverkoilla.....	6
4	TOISTUVAT NEUROVERKOT OSAKKEIDEN HINTOJEN ENNUSTAMISES- SA	8
	4.1 Tutkimusten rakenne	8
	4.2 Tutkimusten tulokset	9
	4.3 Johtopäätökset	11
5	YHTEENVETO.....	13
	LÄHTEET	14

1 Johdanto

Osakkeiden hintojen ennustaminen on ollut osa osakemarkkinoita sen syntymästä lähtien. Sijoittajat ovat pyrkineet saamaan tietoa mitä ostaa ja mitä myydä eri keinoilla läpi historian, mutta vielä ei ole löydetty tapaa, mikä antaisi sijoittajille todellisen etulyöntiaseman muihin tapoihin verrattuna. Osakkeiden hinta muodostuu monista eri tekijöistä, kuten maan poliittisista tilanteesta, markkinoiden yleisestä tilanteesta ja yritysten omasta toiminnasta. Osakkeiden hintaa on täten vaikea ennustaa muuttuvien tekijöiden takia.

Toistuvat neuroverkot pystyvät arvioimaan, mikä funktionaalinen muoto parhaiten kuvastaa aikasarjaa ja ne pärjäsivät perinteisiä aikasarja-analyysejä paremmin 1990-luvulla (Hill, O'Connor ja Remus 1996). Toistuvien neuroverkkojen lupaavan kehityksen takia kiinnostus niitä kohtaan nousi 2000-luvulla ja niiden kykyä aikasarjojen kanssa alettiin tutkia paljon. Samalla mielenkiinto osakesijoittamiseen on noussut runsaasti 2000-luvulla, sen tullessa helpommaksi teknologian avulla ja sen tarjoamat hyvät tuotot riskin ohella on johtanut sijoittajia etsimään uusia tapoja saadakseen voittoja osakemarkkinoilla.

Tutkielma keskittyy tutkimaan kahta toistuvaa neuroverkkoa osakkeiden hinnan ennustamisessa ja arvioimaan, kumpi niistä tarjoaa tarkempia ennusteita. Johdanto pyrkii avaamaan tutkielman taustaa ja sen tavoitteita. Luku 2 sukeltaa syvemmälle neuroverkkojen historiaan, toimintaan ja kahden toistuvan neuroverkon erityispiirteisiin. Osakemarkkinoiden maailma avautuu luvussa 3, jossa tarkastellaan osakemarkkinoita ja osakkeiden ennustamisen haasteita. Luku 4 vertailee kahta toistuvaa neuroverkkoa niiden kyvyssä ennustaa osakkeiden hintoja. Viimeisessä luvussa koostetaan tutkielman tärkeimmät tulokset ja hahmotellaan mahdollisia suuntia tuleville tutkimuksille.

2 Neuroverkot

Fausettin (1994) mukaan neuroverkkojen kehitys alkoi 1940-luvulla, jolloin haluttiin ymmärtää aivoja ja jäljitellä sen vahvuuksia. Nopea edistys tietokoneiden laskennallisessa tehokkuudessa johti siihen, että neuroverkot eivät saaneet vielä paljon huomiota. Lisäksi esioletukset neuroverkkojen kyvyistä oli liioiteltuja, mikä ei auttanut neuroverkkojen saamaa huomiota.

Neuroverkot ovat hermoston mukaan suunniteltu tiedonkäsittelyjärjestelmä, jossa on tehty matemaattisia oletuksia ihmisen kognitiosta. Fausett (1994) kertoo näiden oletuksien olevan:

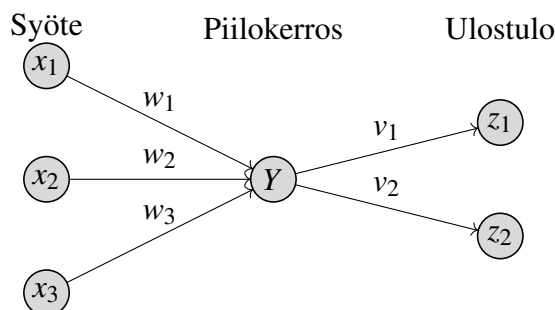
1. Informaation käsittely tapahtuu monessa elementissä, nimeltä neuronit.
2. Signaalit kulkevat neuronien välillä yhteyslinkkien avulla.
3. Jokaisella yhteyslinkillä on oma painonsa.
4. Jokainen neuronit käyttää aktivaatiofunktioitaan painotettujen sisääntulosignaalien summaan saadakseen ulostulosignaalin.

Fausett (1994) selittää kirjassaan neuroverkkojen koostuvan useasta elementistä nimeltä neuronit, jotka ovat yhteyksissä toisiinsa neuroneihin linkillä. Jokaisella linkillä on oma painonsa, jotka kuvastavat verkon käyttämää dataa ratkaistaksensa ongelma.

Lisäksi Fausettin (1994) mukaan jokaisella neuronilla on sisäinen tila (engl. *activity level*), joka on sen syötteiden funktio. Neuronit lähettävät yhden aktivaatio-signaalin kerrallaan usealle neuronille. Vuorostaan nämä neuronit lähettävät signaalin eteenpäin muille neuroneille. Neuronin saama nettosyöte on sen saamien painotettujen syötteiden summa, ja neuronin aktivointi määräytyy sen nettosyötteiden funktiosta. Yleisesti käytetty funktio on logistinen sigmoid funktio.

Eli neuroverkko saa syötteen. Ensiksi lasketaan neuronin saamien syötteiden painotettu summa. Sen jälkeen funktiossa lasketaan summan avulla neuronin aktivaatiotaso. Tämä taas lähetetään mahdollisesti muille neuroneille syötteeksi. Tämä on esitetty kuviossa 1, jossa laskenta tapahtuu neuroverkon piilokerroksessa (engl. *hidden layer*). Neuroverkot koostuvat syötekerroksesta, piilokerroksesta ja ulostulokerroksesta. Piilokerroksesta neuronit aktivoituvat,

mutta aktivoitumista ei voi havaita syötteiden tai ulostulojen kautta. Tästä tulee kerroksen nimi.



Kuvio 1: Yksinkertainen neuroverkko. (Fausett 1994)

$$Y:n \text{ syöte} = w_1x_1 + w_2x_2 + w_3x_3 \quad (2.1)$$

$$w_n = \text{Yhteyden } n \text{ paino} \quad (2.2)$$

$$x_n = \text{Neuronin } n \text{ lähettämä syöte} \quad (2.3)$$

$$\text{Logistinen Sigmoid funktio: } f(x) = \frac{1}{1 + \exp(-x)} \quad (2.4)$$

Neuroverkot pystyvät arvioimaan epälineaarisia funktioita, mikä tekee niistä todella hyödyllisiä aikasarjojen ennustamisessa. (Dang Khoa, Sakakibara ja Nishikawa 2006)

2.1 Toistuvat neuroverkot

Yksinkertaisessa neuroverkossa informaatio etenee syötteestä, piilotettujen yksiköiden kautta ulostuloon. Toistuvat neuroverkot eroavat tästä siten, että ne omaavat takaisinkytkentä yhteyden, joka antaa niille ominaisuuden ”muistaa” neuroverkon aikaisemmat syötteet ja tulot.

Takaisinkytkennän kanssa tulee myös ongelma, nimeltä katoava gradientti (engl. *vanishing gradient*) (Hochreiter 1991). Toistuva neuroverkko on vaikea kouluttaa, sillä jos gradientti muuttuu pieneksi, tai kasvaa suureksi, kun se kulkee takaisin verkossa neuroverkko ei muista aikaisempaa dataa ja keskittyy vain kaikkein tuoreimpaan dataan neuroverkossa. Tämä on todella ongelmallista esimerkiksi, kun koulutetaan toistuvaa neuroverkkoa aikasarjojen ennustamisessa, kuten osakehintojen ennustamisessa.

Seuraavassa osiossa tutkimme toistuvia neuroverkkoja, jotka ovat suunniteltu ratkaisemaan tämä ongelma.

2.2 Pitkä lyhytaikainen muisti (LSTM, engl. *long short-term memory*)

Hochreiter (1991) näytti, kuinka vaikeaa on kouluttaa syviä neuroverkkoja, katoavan gradientti ongelman takia. Tästä seuraten syntyi LSTM, joka on Jürgen Schmidhuberin ja Sepp Hochreiterin vuonna 1997 esittämä menetelmä, suunniteltu minimoimaan tämä ongelma. (Schmidhuber ja Hochreiter 1997)

Schmidhuber ja Hochreiter (1997) kertovat tutkimuksessaan keksimänsä toistuvan neuroverkon arkkitehtuurista ja siitä, kuinka se poikkeaa muista toistuvista neuroverkoista. Heidän mukaansa LSTM:n ero muihin toistuviin neuroverkkoihin perustuu sen sisältämään muistisolun (engl. *memory cell*), joka korvaa perinteisen neuronin, sekä muistosolun sisältämien porttiyksikköjen (eng. *gate units*) käyttöön. Sisääntuloportti suojaa muistosolun sisältöä turhilta tiedoilta, kun taas ulostuloportti suojaa muita soluja tarpeettomalta tiedolta. Tämä rakenteellinen ero ehkäisee katoavan gradientti ongelman.

Schmidhuber ja Hochreiterin esittämä menetelmä ei silti onnistunut oppimaan kunnolla pitkissä aikasarjoissa. Tämän takia kaksi vuotta myöhemmin Schmidhuber, Gers ja Cummins (1999) esittivät ratkaisun LSTM:n ongelmaan. Heidän esittämä unohdusportti (engl. *forget gate*) antoi LSTM neuroverkolle kyvyn nollata muistisolun, kun niiden sisältö on vanhentunut.

2.3 Aidatut toistuvat yksiköt (GRUs, engl. *gated recurrent units*)

Zhang et al. (2020) selittävät kirjassaan, kuinka LSTM:n suosion kasvaessa 2010-luvulla tutkijat pyrkivät muokkaamaan sen arkkitehtuuria. Tavoitteena oli säilyttää LSTM:n perusideat, kuten porttimekanismit, samalla etsien laskennallisesti nopeampaa lähestymistapaa.

Lisäksi Zhang et al. (2020) kertovat kuinka Cho et al. esittelivät vuonna 2014 GRU-menetelmän, joka tarjosi samantasoisen suorituskyvyn, mutta oli samalla laskennallisesti nopeampi kuin LSTM.

Zhang et al. (2020) mukaan Cho et al. esittämä GRU korvaa LSTM:n kolmet portit nol-
lausportilla (engl. *reset gate*) ja päivitysportilla (engl. *update gate*). Nollausportti kontrolloi,
kuinka paljon edellisestä tilasta halutaan muistaa ja päivitysportti taas kontrolloi soluun päi-
vitettävän tiedon määrää. Eli nollaus portti auttaa sekvenssien lyhytaikaisissa riippuvuuksis-
sa ja päivitysportti pitkäaikaisissa riippuvuuksissa.

Seuraavassa osiossa tutkimme osakemarkkinoita ja voiko osakkeita ennustaa.

3 Osakemarkkinat

Tutkielmassa keskitymme vertailemaan neuroverkkojen suorituskkyä osakkeiden hintojen ennustamisessa. Tämän vuoksi on tärkeää taustoittaa neuroverkot ja tutkielmassa käsiteltävät arkkitehtuurit. Samalla on hyvä tietää osakemarkkinoista ja niihin liittyvistä käsitteistä, sillä tässä kontekstissa neuroverkot tulevat toimimaan.

3.1 Yritys ja osakkeet

Osakkeenomistajalla on oikeuksia siihen yritykseen, minkä osakkeen hän omistaa. Näihin sisältyy äänioikeus yhtiökokouksessa, oikeus osinkoihin, jos yhtiökokous päättää jakaa voitot osinkoina, mutta tarkemmin sijoittajalla on oikeus näihin voittoihin, mikä voi näkyä myös osakkeen hinnan nousuna.

Osakkeet ovat listattuna pörssissä, josta sijoittavat voivat ostaa ja myydä osakkeita, kun myyjän ja ostajan hinnat kohtaavat. Osakkeet päätyvät pörssiin myyntiin, kun yritys päättää myydä osakkeitansa ja listautua pörssiin, samalla kasvattaen yrityksen pääomaa IPO:ssa (*initial public offering*). Täten yritystä koskee myös uudet velvoitteet, kuten tiedonantovelvoite, jonka mukaan yrityksen tulee tiedottaa kaikki relevantti tieto välittömästi. Tämä on tärkeä osa osakkeen hinnan määrittelyä, sillä tehokkaiden markkinoiden hypoteesin (Malkiel ja Fama 1970) mukaan osakkeen hinnassa on kaikki tiedossa oleva tieto. Tämän huomaa, kun yritys ilmoittaa hyvistä, tai huonoista uutisista. Tätä seuraa saman tien markkinoiden reaktio osakkeen hinnassa.

3.2 Osakkeiden ennustaminen neuroverkoilla

Tehokkaiden markkinoiden hypoteesin (Malkiel ja Fama 1970) mukaan markkinoiden toimiessa tehokkaasti ja sijoittajien ollessa rationaalisia osakkeen markkinahinnan tulisi olla sama kuin osakkeen fundamenttiarvo. Osakkeen fundamenttiarvo tarkoittaa liiketoiminnan rahavirroista määräytyvää osakkeen arvoa. Tämä eroaa kuitenkin todellisuudesta, jossa tunteet, odotukset ja muut tekijät vaikuttavat osto ja myynti päätöksiin. Tästä johtuen hinnat

voivat vaihdella fundamenttiarvosta ja markkinat voivat olla epätehokkaita.

Viisi yleisesti käytettyä indikaattoria, joita käytetään, kun koulutetaan neuroverkkoa osakkeen hinnan ennustamiseen, ovat osakkeen avaushinta, osakkeen sulkuhinta, osakkeen päivän korkein hinta, osakkeen päivän alin hinta ja päivän aikana myytyt tai ostetut osakkeet. Myös muita teknisiä indikaattoreita voidaan käyttää, mutta nämä ovat yleisimmät.

Gao et al. (2021) mukaan osakkeiden hinnat vaihtelevat usein erilaisista syistä, jotka johtavat epälineaarisiin suhteisiin osakkeiden tiedoissa. Epälineaariset suhteet viittaavat siihen, että muutos yhdessä muuttujassa, ei ole suorassa yhteydessä muutoksiin muissa muuttujissa. Esimerkiksi muutos päivän aikana myytyissä osakkeissa, ei suoraan kerro nouseeko vai laskeeko osakkeen hinta.

Neuroverkot ei tietenkään voi täydellisesti ennustaa osakkeiden hintaa. Jos se olisi mahdollista voisi sijoittajat saada loputonta voittoa ja tulla äärettömän rikkaiksi, mutta neuroverkot voivat oppia datassa olevia trendejä ja tehdä ennustuksia osakkeen tulevaisuuden vaihteluista, toisin sanoen, olla suuntaa antavia. Lisäksi, jos markkinat toimisivat aina täydellisen tehokkaasti olisi osakkeiden ennustaminen neuroverkkojen avulla vaikeaa, ellei jopa mahdotonta. Tällöin osakkeen arvossa olisi kaikki relevantti tieto, jolloin trendejä osakkeen hinnassa olisi vaikeampi huomata.

LSTM- ja GRU mallien suoriutumiskyky aikasarjojen kanssa on luontoisesti johtanut tutkimuksiin näiden mallien kyvyistä osakkeiden hintojen ennustamisessa. Seuraavassa osiossa tarkastellaan näitä tutkimuksia ja mallien suoriutumista.

4 Toistuvat neuroverkot osakkeiden hintojen ennustamisessa

LSTM:n ja GRU:n hyvä suoriutuminen aikasarjojen ennustamisessa on johtanut tutkimuksiin niiden kyvyistä osakkeiden hintojen ennustamisessa. Tässä osiossa ensiksi kerron yleisesti tutkimusten rakenteista, sillä prosessit tutkimuksissa ovat pitkälti samat. Sen jälkeen tutkimme muutamaa uudempaa tutkimusta ja niiden tuloksia. On myös hyvä muistaa, että tämä katsaus on rajoitettu näihin kahteen neuroverkkomalliin tutkielman selkeyden ja laajuuden vuoksi. Laajempi katsaus mahdollistaisi useamman mallin tarkastelun, mikä voisi antaa monipuolisemman kuvan toistuvien neuroverkkojen suoriutumisesta osakkeiden hinnan ennustamisessa.

4.1 Tutkimusten rakenne

Kaikki tutkimukset, joita tulemme tarkastelemaan seuraavat pitkälti seuraavaa menetelmää. Ne lataavat julkisesta rajapinnasta, kuten Yahoo! Financesta osakkeen historian tietyltä aikaväliltä. Tämän jälkeen tutkimuksissa käytetään Pythonin avoimia ohjelmointikirjastoja, kuten TensorFlow tai PyTorchia toistuvan neuroverkko mallin kouluttamiseen. Taulukossa 1 on esimerkki Yahoo Financesta saaduista Alphabetin osaketiedoista päivältä 9.10.2009. Yleisesti käytetty tapa on myös ottaa valitsemasta datasta osa, esim. 80% mallin kouluttamiseen ja loput 20% käytetään testaamaan mallin tarkkuus. Eli katsotaan, kuinka lähelle oikeita hintoja malli ennustaa.

Date	Open	High	Low	Close	Adj Close	Volume
Oct 09, 2009	12.87	12.99	12.81	12.86	12.86	109,934,523

Taulukko 1: Alphabetin osakkeen tiedot 9.10.2009, valuutta USD (Yahoo-Finance)

Näistä sarakkeista usein valitaan ”close”, joka tarkoittaa osakkeen sulkuhintaa, tai ”adjusted close”, jota neuroverkolla halutaan ennustaa. ”Adjusted close” sisältää hinnassaan osakesplit ja osingot. Tutkimuksissa käytetyt tekniset indikaattorit usein vaihtelee, mutta taulukossa 1

esitetty osakkeen tiedot ovat yleensä mukana. Sarakkeista "open" kuvaa sen päivän ensimmäisen osakkeen ostohintaa, "high" taas päivän korkeinta hintaa, "low" matalinta hintaa ja "volume" päivän aikana myytyjen osakkeiden määrää.

Tutkimuksissa mallien tarkkuutta arvioidaan myös pitkälti samoilla menetelmillä. Yleisesti käytettyjä tapoja ovat *the mean square error* (MSE), *root mean square error* (RMSE), *mean absolute error* (MAE) ja *coefficient of determination* (R^2)

Edellä mainitut menetelmät painottavat ja mittaavat eri arvoja. Tätä tutkielmaa varten tarvitsee vain tietää, että MSE, MAE ja RMSE mittaavat virheitä ja niiden minimointi on hyvä, kun taas R^2 arvioi mallin sovittamista dataan ja sen maksimointi on hyvä. Yhdessä nämä menetelmät kuvaavat hyvin neuroverkko mallin suorituskykyä aikasarjojen ennustamisessa.

4.2 Tutkimusten tulokset

Ensimmäinen tutkimus, jota tarkastelemme, on luonut Jiayu Qiu et al. vuonna 2020. (Qiu, Wang ja Zhou 2020)

Tutkimuksessa he käyttivät kolmea osakeindeksiä neuroverkkojen kouluttamiseen. Heidän valitsemansa indeksit olivat S&P 500 Index (S&P 500), Dow Jones Industrial Average (DJIA) ja Hang Seng Index (HSI). He käyttivät tutkimuksessaan, sekä LSTM- että GRU-malleja, jonka takia tämä tutkielma sopii hyvin tarkastelun kohteeksi. He käyttivät lisäksi kahta muuta neuroverkko mallia, mutta jätämme ne tämän tarkastelun ulkopuolelle ja keskitymme pelkästään tarkastelemaan heidän tavallisia LSTM ja GRU malleja. Lisäksi he ennustivat indeksien avaushintaa, joka taulukossa 1 olisi "open" sarake.

DJIA				
Malli	MAE	MSE	RMSE	R^2
LSTM	0.2360	0.0785	0.2802	0.9235
GRU	0.1737	0.0452	0.2126	0.9560

Taulukko 2: LSTM- ja GRU-mallien vertailu DJIA-datasetissä (Qiu, Wang ja Zhou 2020)

S&P 500				
Malli	MAE	MSE	RMSE	R²
LSTM	0.2676	0.1208	0.3475	0.8829
GRU	0.2394	0.1000	0.3162	0.9030

Taulukko 3: LSTM- ja GRU-mallien vertailu S&P 500 -datasetissä (Qiu, Wang ja Zhou 2020)

HSI				
Malli	MAE	MSE	RMSE	R²
LSTM	0.3031	0.1839	0.4288	0.8097
GRU	0.2659	0.1409	0.3754	0.8541

Taulukko 4: LSTM- ja GRU-mallien vertailu HSI-datasetissä (Qiu, Wang ja Zhou 2020)

Tutkimuksen tulosten mukaan GRU-malli pärjasi LSTM-mallia paremmin näiden kolmen indeksien ennustamisessa. Tämä eroaa Sethian ja Rautin (Sethia ja Raut 2019) tutkimuksen tuloksista, jossa he käyttivät 50 teknistä indikaattoria kouluttamaan neuroverkkoja ennustamaan S&P 500-indeksiä

Tutkielmassa LSTM-mallilla oli korkeampi R² ja matalampi RMSE, kuin GRU-mallilla. Tutkimuksen tulokset näkyvät taulukossa 5.

Malli	RMSE	R²
LSTM	0.000428	0.948616
GRU	0.000511	0.938098

Taulukko 5: LSTM- ja GRU-mallien vertailu S&P 500 -datasetissä (Sethia ja Raut 2019)

Jos vertailemme molempien tutkimusten mallien suoriutumista S&P 500 -datasetissä Sethian ja Rautin malli on pärjännyt paljon paremmin. Molemmissa tutkimuksissa on käytetty melko samalta aikaväliltä dataa. Qiu et al. ottivat datan 3.1.2000 – 16.5.2019 väliltä, kun taas Sethia ja Raut ottivat datan 3.1.2000 - 30.10.2017 väliltä. Suurin tekijä tälle erolle on tutkimuksissa

käytettyjen teknisten indikaattorien määrä. Sethia ja Raut käyttivät yli 50 eri indikaattoria, kun taas Qiu et al. käyttivät kuutta tavallista muuttujaa, jotka ovat esitelty kuviossa 1. Lisäksi Sethian ja Rautin tutkimuksessa on käytetty suurempaa testausdataa, mikä on voinut myös antaa etulyöntiaseman LSTM mallille.

On epäselvää miksi Qiu et al. tutkimuksessa GRU-malli suoriutui paremmin suurien aikasarjojen ennustamisessa, kuin LSTM-malli, joka yleensä pärjää paremmin suurien aikasarjojen kanssa (Yang, Yu ja Zhou 2020). He ei eivät kerro käyttämistään LSTM ja GRU neuroverkkoista paljoa, sillä tutkimus keskittyi heidän esittämään neuroverkko mallin, joten voi vain spekuloida, miksi GRU pärjasi paremmin. Todennäköisesti tähän vaikutti useat eri tekijät, kuten heidän valitsemansa data mallinkoulutukseen, testaukseen ja muut datan esikäsitteilyt. Esimerkiksi Sethian ja Rautin tutkimuksessa käytetty testausdata oli ajalta 12.6.2014 - 23.10.2017, kun taas Qiu et al. tutkimuksen indekseille se oli ajalta 17.5.2019 - 1.7.2019.

Seuraavassa osiossa käydään läpi johtopäätökset näiden tutkimusten pohjalta.

4.3 Johtopäätökset

On vaikea antaa yksiselitteistä vastausta siihen, kumpi neuroverkkomalli on parempi osakkeiden ennustamisessa, sillä tutkimustulokset vaihtelevat. Ensimmäisessä tutkimuksessa havaittiin, että GRU pärjasi paremmin, kun taas toisessa tutkimuksessa LSTM osoittautui tehokkaammaksi. Gao et al. (2021) tutkimuksen tuloksien mukaan sekä GRU että LSTM kykenevät tehokkaasti ennustamaan osakkeita, eikä kumpikaan osoittautunut toista paremmaksi. Neuroverkkojen suorituskykyyn vaikuttaa moni tekijä, kuten valittu aikasarja, käytetty testi- ja koulutusdata sekä neuroverkon konfiguraatio, kuten piilokerroksen neuronien määrä, kerrosten lukumäärä, oppimisnopeus ja aktivointifunktion tyyppi.

Sethian ja Rautin tutkimuksessa, jossa käytettiin enemmän teknisiä indikaattoreita ja pidempää testidataa, havaittiin LSTM-mallin pärjäävän paremmin kuin GRU. Tämä tulokseen vaikuttava tekijä ei ole yllättävä, sillä teknisten indikaattorien runsaampi määrä ja pidempi testijakso antavat LSTM:lle mahdollisuuden oppia monimutkaisempia suhteita ja trendejä.

Saudin ja Shakyen tutkimuksessa (2020) GRU pärjasi LSTM:ää paremmin osakkeiden hin-

tojen ennustamisessa. He huomauttavat, että GRU:n parempi tarkkuus johtui osassa kohtalaisesta datamäärästä, jolla malli koulutettiin. Eli toisin sanoen, jos mallin koulutukseen olisi ollut enemmän dataa, olisi LSTM voinut pärjätä GRU:ta paremmin.

Samanlaisen tulokseen tuli Weiss et al. tutkimus (2018), jossa GRU myös voitti LSTM:än. Tutkimuksessaan he suosittelivat GRU:n käyttöä LSTM sijasta aikasarjojen ennustamisessa.

Osassa tarkastelluissa tutkimuksissa tutkijoilla on oma malli, jonka he tarjoavat vaihtoehtona GRU:lle ja LSTM:lle. Esimerkiksi Qiu et al (2020) esittivät oman mallinsa nimeltä WLSTM+Attention (engl. *wavelet transform long short-term memory with attention mechanism*). Heidän esittämänsä malli pärjasi paljon paremmin verrattuna tavalliseen LSTM:ään.

S&P 500				
Malli	MAE	MSE	RMSE	R²
LSTM	0.2676	0.1208	0.3475	0.8829
WLSTM+Attention	0.1935	0.0546	0.2337	0.9470

Taulukko 6: LSTM- ja Wavelet Transform Long Short-Term Memory with Attention Mechanism-mallien vertailu S&P 500 -datasetissä (Qiu, Wang ja Zhou 2020)

Tarkastelujen tutkimusten perusteella voidaan päätellä, että joko GRU tai LSTM voi suoriutua paremmin tilanteesta riippuen. Lisäksi näitä malleja voi käyttää lähtökohtana, ja lisäämällä niihin erilaisia ominaisuuksia voi saavuttaa parempia tuloksia kuin perinteisillä GRU- tai LSTM-malleilla.

Kuten tämän osion alussa mainittiin, tarkastelu ei ole laajamittainen ja tutkimus voisi tulla vahvempiin johtopäätöksiin, jos olisi tarkasteltu enemmän alan tutkimuksia.

5 Yhteenveto

Tässä tutkielmassa tutkittiin kahden toistuvan neuroverkon suoriutumista osakehintojen ennustamisessa. Tutkimusten tulokset vaihtelivat, eikä ole yksiselitteistä vastausta siihen, kumpi neuroverkoista tarjoaa tarkempia ennusteita. Tutkimukset saivat erillaisia tuloksia ja osassa tutkimuksissa tutkijoiden omat mallit saivat parempia tuloksia, kuin pelkkä LSTM tai GRU.

Osiossa 4.3 listattiin syitä, jotka vaikeuttavat vahvan tulokseen saapumista. Näihin kuuluivat muun muassa tutkittujen osakkeiden vaihtelu, aikaväli, tekniset indikaattorit sekä koulutus- ja testidata. Näiden ja muiden tekijöiden perusteella joko LSTM- tai GRU-malli voi tuottaa parempia tuloksia.

Yhteenvetona voidaan todeta, että jos käytössä on runsaasti teknisiä indikaattoreita ja laaja aikasarja, voi LSTM-malli suorittaa paremmin. Toisaalta, jos käytettävissä on kohtalainen määrä dataa ja halutaan laskennallisesti nopeampi malli, silloin GRU-malli voi olla parempi vaihtoehto.

Tutkielman tulosten valossa olisi seuraavaksi hyvä vertailla GRU- ja LSTM- mallien päälle luotujen mallien, kuten osiossa 4.3 mainitun WLSTM+Attention, suorituskykyä osakehintojen ennustamisessa, sillä niiden suorituskyky näyttää voittavan tavalliset LSTM- ja GRU-mallit.

Lähteet

Dang Khoa, Nguyen Lu, Kazutoshi Sakakibara ja Ikuko Nishikawa. 2006. “Stock Price Forecasting using Back Propagation Neural Networks with Time and Profit Based Adjusted Weight Factors”. Teoksessa *2006 SICE-ICASE International Joint Conference*, 5484–5488. <https://doi.org/10.1109/SICE.2006.315683>.

Fausett, Laurene V. 1994. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall, Inc. ISBN: 978-0-13-334186-7.

Gao, Ya, Rong Wang ja Enmin Zhou. 2021. “Stock Prediction Based on Optimized LSTM and GRU Models”. *Scientific Programming* 2021:4055281. <https://doi.org/10.1155/2021/4055281>.

Gers, F.A., J. Schmidhuber ja F. Cummins. 1999. “Learning to forget: continual prediction with LSTM”. Teoksessa *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, nide 2, 850–855 vol.2. <https://doi.org/10.1049/cp:19991218>.

Hill, Theodore, Marcus O’Connor ja William Remus. 1996. “Neural Network Models for Time Series Forecasts”. *Management Science* 42 (7): 1082–1092. <http://www.jstor.org/stable/2634369>.

Hochreiter, Sepp. 1991. “Untersuchungen zu dynamischen neuronalen Netzen”. *Diploma thesis*.

Malkiel, Burton G. ja Eugene F. Fama. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work”. *The Journal of Finance* 25 (2): 383–417.

Qiu, Jie, Bo Wang ja Can Zhou. 2020. “Forecasting stock prices with long-short term memory neural network based on attention mechanism”. *PLoS ONE* 15 (1): e0227222. <https://doi.org/10.1371/journal.pone.0227222>. <https://doi.org/10.1371/journal.pone.0227222>.

- Saud, Arjun Singh ja Subarna Shakya. 2020. “Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE”. International Conference on Computational Intelligence and Data Science, *Procedia Computer Science* 167:788–798. ISSN: 1877-0509. <https://doi.org/https://doi.org/10.1016/j.procs.2020.03.419>. <https://www.sciencedirect.com/science/article/pii/S1877050920308851>.
- Schmidhuber, Jürgen ja Sepp Hochreiter. 1997. “Long Short-Term Memory”. *Neural Computation* 9 (8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Sethia, Akhil ja Purva Raut. 2019. “Application of LSTM, GRU and ICA for Stock Price Prediction: Proceedings of ICTIS 2018, Volume 2”, 479–487. Tammikuu. ISBN: 978-981-13-1746-0. https://doi.org/10.1007/978-981-13-1747-7_46.
- Weiss, Gail, Yoav Goldberg ja Eran Yahav. 2018. *On the Practical Computational Power of Finite Precision RNNs for Language Recognition*. arXiv: 1805.04908 [cs.LG].
- Yahoo-Finance. *Alphabetin osakeen tiedot 9.10.2009*. <https://finance.yahoo.com/quote/GOOGL/history?p=GOOGL>. Haettu päivämäärällä 9.10.2009.
- Yang, Shudong, Xueying Yu ja Ying Zhou. 2020. “LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example”. Teoksessa *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, 98–101. <https://doi.org/10.1109/IWECAI50956.2020.00027>.
- Zhang, Aston, Zachary C. Lipton, Mu Li ja Alexander J. Smola. 2020. *Dive into Deep Learning*. d2l.ai. <https://d2l.ai/d2l-en.pdf>.