



TALLINNA TEHNIKAÜLIKOOL
INSENERITEADUSKOND
Virumaa kolledž

**Aegridade prognoosimine Power BI tarkvara abil
aktsiahindade näitel**

**Time Series Forecasting on Stock Prices Using Power BI
Software**

TELEMAATIKA JA ARUKATE SÜSTEEMIDE ÕPPEKAVA LÕPUTÖÖ

Üliõpilane: Lutšika Klim

Üliõpilaskood: 182740EDTR

Juhendaja: Olga Dunajeva, vanemlektor

AUTORIDEKLARATSIOON

Olen koostanud lõputöö iseseisvalt.

Lõputöö alusel ei ole varem kutse- või teaduskraadi või inseneridiplomit taotletud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

"...." 20.....

Autor:

/ allkiri /

Töö vastab rakenduskõrgharidusõppe lõputööle/magistritööle esitatud nõuetele

"...." 20.....

Juhendaja:

/ allkiri /

Kaitsmisele lubatud

"...." 20.....

Kaitsmiskomisjoni esimees

/ nimi ja allkiri /

LIHTLITSENTS LÕPUTÖÖ ÜLDSUSELE KÄTTESAADAVAKS TEGEMISEKS JA REPRODUTSEERIMISEKS

Mina Lutšika Klim (sünnikuupäev: 25.12.1985)

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Aegridade prognoosimine Power BI tarkvara abil aktsiahindade näitel“, mille juhendaja on Olga Dunajeva,
 - 1.1. reprodutseerimiseks säilitamise ja elektroonilise avaldamise eesmärgil, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta kolmandate isikute intellektuaalomandi ega isikuandmete kaitse seadusest ja teistest õigusaktidest tulenevaid õigusi.

SISUKORD

EESSÕNA	6
SISSEJUHATUS	7
1 TEOREETILINE ÜLEVAADE	8
1.1 Aegridade prognoosimise meetodid	8
1.1.1 ARIMA mudel.....	8
1.1.2 Prophet mudel	9
1.1.3 Närvivõrgu mudelid	9
1.1.4 Mudelite valideerimine	10
2 MEETODID JA VAHENDID	11
2.1 Andmete kirjeldus	11
2.2 Aegridade prognoosimise vahendid POWER BI-s	12
2.2.1 Power BI visuaalid aegridade prognoosimiseks.....	13
2.2.2 Masinõppe vahendid Power BI-s	16
3 TÖÖ TULEMUSED	18
3.1 Andmete eeltöötlus.....	18
3.2 Modelleerimine Power BI visuaalide abil.....	21
3.3 Automaatmodelleerimine Azure ML Studios	24
3.3.1 Modelleerimise protsess	24
3.3.2 Azure mudelite rakendamine Power BI-s.....	30
3.4 Modelleerimine Pythoni abil.....	33
3.4.1 Pythoni skriptid Power BI-s	33
3.4.2 Pythoni mudelite rakendamine Power BI-s	33
KOKKUVÕTE	36
SUMMARY	37
KASUTATUD KIRJANDUSE LOETELU	39

EESSÕNA

Antud lõputöö on koostatud ja sõnastatud töö autori Lutšika Klimi ja juhendaja Olga Dunajeva koostöös. Teema valik oli juhendaja poolt pakutud ning on tingitud autori huvist aktsiaturgude vastu ja soovist edasi arendada oma seniseid teadmiseid Power BI tarkvaraga, lähtudes ka Tallinna Tehnikaülikooli Virumaa kolledži vajadusest Power BI õppematerjali järele.

Käesolevas lõputöös autor valis modelleerimiseks ja prognoosimiseks Netflix aktsiahindade andmestiku Yahoo.Finace portaalist [1].

Autor soovib tänada juhendaja Olga Dunajevat abi eest lõputöö teema valikul, lõputöö vormistamisel, konsultatsioonide ja juhendamise eest.

Võtmesõnad: Power BI, Azure ML, aegridade modelleerimine, masinõpe, prognoosimise mudelid, diplomitöö.

SISSEJUHATUS

Alates 17. sajandist investorid on kasutanud turusündmuste ennustamiseks tehnilist analüüsi [2]. Selline meetod sobib hästi lühiajaliste strateegiate jaoks ja põhineb minevikus toimunud hinnamuutuste mustrite uurimisel. Kuid aga masinõppe ja erinevate BI (*Business Intelligence*) tarkvara arenemisega on nüüd võimalus aktsiahindasid analüüsida ja prognoosida ka automaatrežiimis. Tänapäeva maailmas, kus investeerimine ei ole enam ainult rikastele inimestele, üsna rohkem tekkib huvi ja nõudlus erinevate aktsiahindade prognoosimise meetodite vastu, mis aitavad investoril koostada oma investeerimisplaani.

Lõputöö teema valik on tingitud autori huvist aktsiaturgude vastu. Aktsiahindade modelleerimise ja prognoosimine Power BI tarkvara abil sai valitud, sellepärast, et autoril on juba eelnevalt olnud kokkupuuteid selle tarkvaraga ning oli soov ka oma teadmised edasi arendada.

Käesoleva lõputöö põhieesmärgiks on uurida Power BI aegridade prognoosimise ja masinõppe vahendeid aktsiahindade modelleerimise näitel. Töös vaadeldakse erinevaid aegridade prognoosimise ja edasise käitumise modelleerimise meetodeid ja masinõppe mudeleid, mida Power BI tarkvara vahendid pakuvad. Tehakse nende mudelite täpsuste võrdlusanalüüsi aktsiahindade andmete näitel ning pakutakse aktsiahindade ja nende prognooside visualisatsioonid. Lõputöö eesmärgist lähtuvalt on püstitatud ka lõputöö ülesanded: uurida Power BI-sse sisseehitatud masinõppemeetodid, mis lubavad aegridasid prognoosida; konstrueerida nende abil mudelid aktsiahindade modelleerimiseks Yahoo.Finace portaalist saadud andmete põhjal ning koostada mudelite konstrueerimise protsessist ülevaade; võrrelda saadud mudelite täpsused ning teha parimate mudelite põhjal aktsiahindade prognooside visualiseerimine.

Töö koosneb kolmest peatükist. Esimeses peatükis tutvustatakse aegridade prognoosimise meetodeid ja esitatakse kasutatud valideerimise mõõdikut. Teises peatükis on kirjeldatud töös kasutatud andmeid ning Power BI tarkvara aegridade prognoosimise võimalused. Kolmandas peatükis modelleeritakse aktsiahinnad Power BI aegridade prognoosimise meetodite abil, kasutades ka Azure masinõppe integreerimist ja Pytoni skripte ning tehakse parimate mudelite visualisatsioonid.

1 TEOREETILINE ÜLEVAADE

1.1 Aegridade prognoosimise meetodid

Aegrida y_t on ajas järjestatud andmete jada, mis salvestatakse kindla ajavahemiku jooksul. Aegridade prognoosimine on masinõppe tehnika, mis analüüsib andmeid ja ajavahemikke tulevaste sündmuste ennustamiseks. See meetod annab ajalooliste aegridade andmete põhjal peaaegu täpsed eeldused tulevikutrendide kohta. [3]

Aegridade abil saab analüüsida andmete peamisi mustreid, nagu trendid, sesoonsus, tsüklilisus ja ebaregulaarsus. [4]

Aegread koosnevad mitmest komponendist:

- Trend: pikema aja jooksul ilmnev andmete suurenemise või vähenemise tendents.
- Sesoonne komponent: perioodiliselt (iga aasta, või iga kuu, või iga nädal) korduvad muutused trendi suunas.
- Tsükliline komponent: aastast pikema perioodiga perioodilised muutused trendi suunas.
- Juhuslik komponent ehk müra, mis on põhjustatud paljude tegurite koosmõjust ja ei ole prognoositav. [3]

Aegridade prognoosimisel on palju kasutusvõimalusi paljudes tööstusharudes ja ettevõtetes. See võimaldab ennustada nõudlust toodete järele, ennustada müügist saadavat raha või prognoosida aktsiahindade tulevasi väärtusi.

1.1.1 ARIMA mudel

ARIMA mudel - autoregressiivsed integreeritud libiseva keskmisega (*AutoRegressive Integrated Moving Average, ARIMA*) mudelid, mis kuuluvad lineaarsete mudelite hulka ning lubavad kirjeldada nii statsionaarseid (aegridu, mille omadused ei sõltu aegrea vaatlemise ajast) kui ka mittestatsionaarseid aegridu. See on kombinatsioon autoregressiivsest (AR) ja libiseva keskmise (MR) mudelist. AR mudeli prognoos vastab muutuja varasemate väärtuste ehk viitaegade (inglise keeles *lags*) y_{t-1}, \dots, y_{t-p} lineaarkombinatsioonile. Libiseva keskmise mudeli prognoos vastab varasemate prognoosivigade ehk juhuslike komponentide lineaarkombinatsioonile $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$. "I" tähistab andmeväärtuste asendamist nende väärtuste ja eelmiste väärtuste erinevusega (aegrea muudud ehk diferentsid, $\Delta y_t = y_t - y_{t-1}$). [5]

ARIMA mudeleid liigitatakse kolme teguri järgi kasutades tähistust ARIMA(p,d,q), kus

p = autoregressiivse osa järk (AR)

d = diferentside järk – kas andmete asemel kasutatakse aegrea muudud või muutude muudud jne üldjuhul $\Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}$ (I)

q = libiseva keskmise järk (MA) [3]

SARIMA mudel – SARIMA (*Seasonal Autoregressive Integrated Moving Average*) tähistab sesoonsset autoregressiivset integreeritud libiseva keskmisega mudelit. See laiendab ARIMA mudelit, lisades mudelisse sesoonsuse perioodile vastavate nihetega autoregressiivseid ja libiseva keskmise liikmeid kasutades tähistust $ARIMA(p,d,q)(P,D,Q)_m$, kus P, D, Q on järkude p, d, q sesoonsed analoogid, kuid tuleb arvesse võtta sesoonsset nihet y_{t-m} , ja m on sesoonsuse perioodi ajasammude arv. [6]

ARIMA ja SARIMA mudelid on aegridade prognoosimisel kaks kõige laialdasemalt kasutatavat lähenemisviisi. ARIMA mudelite eesmärk on kirjeldada andmetes esinevaid autokorrelatsioone ehk seosi y_t ja y_{t-1} vahel, y_t ja y_{t-2} vahel jne. Seda mudelit kasutatakse nõudluse ja varude, eelarvestamise, müügikvootide, turunduskampaaniate ja hangete ennustamiseks. [3]

1.1.2 Prophet mudel

Prophet on aegridade prognoosimise protseduur, mis põhineb aditiivsel mudelil, kus mittelineaarsed trendid prognoositakse aastase, nädala ja igapäevase sesoonsusega ning arvestades puhkepäevade mõju. See töötab kõige paremini aegridadega, millel on tugev hooajaline mõju ja mitme hooaja ajaloolised andmed. Prophet on vastupidav andmete puudumise ja trendis muutuste suhtes ning saab tavaliselt hästi hakkama andmetes esinevate kõrvalekalletega. [7, 8]

Prophet on avatud lähtekoodiga tarkvara, mille on välja andnud Facebooki Core Data Science meeskond. See on allalaadimiseks saadaval nii R-i kui ka Pythoni programmeerimiskeelte kasutajale. [7, 8]

1.1.3 Närvivõrgu mudelid

Tehisnärvivõrgud on prognoosimismeetodid, mis põhinevad lihtsatel aju matemaatilistel mudelitel. Need võimaldavad leida keerulisi mittelineaarseid seoseid prognoositava muutuja ja selle ennustajate vahel. Süvaõppe närvivõrgud suudavad automaatselt õppida suvalisi keerulisi suhteid sisendmuutujate ja prognoositava muutuja vahel, mistõttu võivad olla aegridade prognoosimisel kasulikud. [8]

LSTM (*Long Short-Term Memory*) - rekurrentne närvivõrk (närvivõrk, mis suudab kasutada oma sisemist olekut ehk mälu, et töödelda jada kujul esitatud informatsiooni), mis tegeleb pikaajaliste sõltuvustega. LSTM kasutatakse laialdaselt aegridade prognoosimisel. See mäletab teavet varasematest andmetest ja on võimeline õppima järjestuse sõltuvust jada olekute ennustamise probleemides. [5]

LSTM lahendab rekurrentse närvivõrgu probleemi: lühike mälu (*short-memory*). Kasutades rida "väravaid" (*gates*), millest igaühel on oma rekurrentne närvivõrk, ning põhinedes tõenäosusmudelil LSTM suudab andmepunkte säilitada, unustada või ignoreerida. [9]

1.1.4 Mudelite valideerimine

Prognoosimise viga on erinevus aegrea tegeliku väärtuse ja selle prognoosi vahel. Aegride prognoosimise täpsuse hindamiseks käesolevas töös on kasutatud prognoosimise vigade põhjal leitavaid mõõdikut - ruutkeskmist viga (*Root Mean Square Error, RMSE*). [3]

RMSE on ruutjuur vigade ruutude keskmisest ja see mõõdik annab ülevaate ennustuse täpsuse hajuvuse või varieeruvuse kohta. RMSE rõhutab suurte vigade olemasolu mudelis, mistõttu RMSE on soovitatav kasutada juhul, kui on soov suuremaid vigu vältida. [3]

$$\text{Root mean squared error: RMSE} = \sqrt{\text{mean}(e_t^2)}$$

Joonis 1.1. RMSE leidmise valem [3]

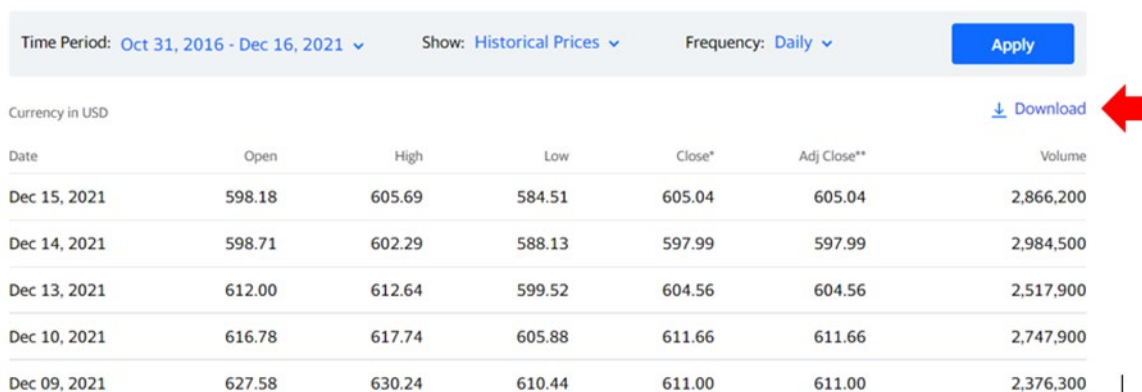
Joonisel 1.1. on näidatud RMSE leidmise valem, kus e_t on aegrea prognoosimise vead.

2 MEETODID JA VAHENDID

2.1 Andmete kirjeldus

Netflix (NFLX) aktsiahindade andmestik, prognoosimiseks ja masinõppe mudelite ehitamiseks on pärit interneti portaalist Yahoo.Finance

<https://finance.yahoo.com/quote/NFLX/history?p=NFLX>, kust saab võtta ajaloolised andmed, valida perioodi, ning alla laadida csv -formaadis faili andmetega (vt Joonis 2.1). Allalaaditud andmestik koosneb 1260 reast ja 7 muutujast.



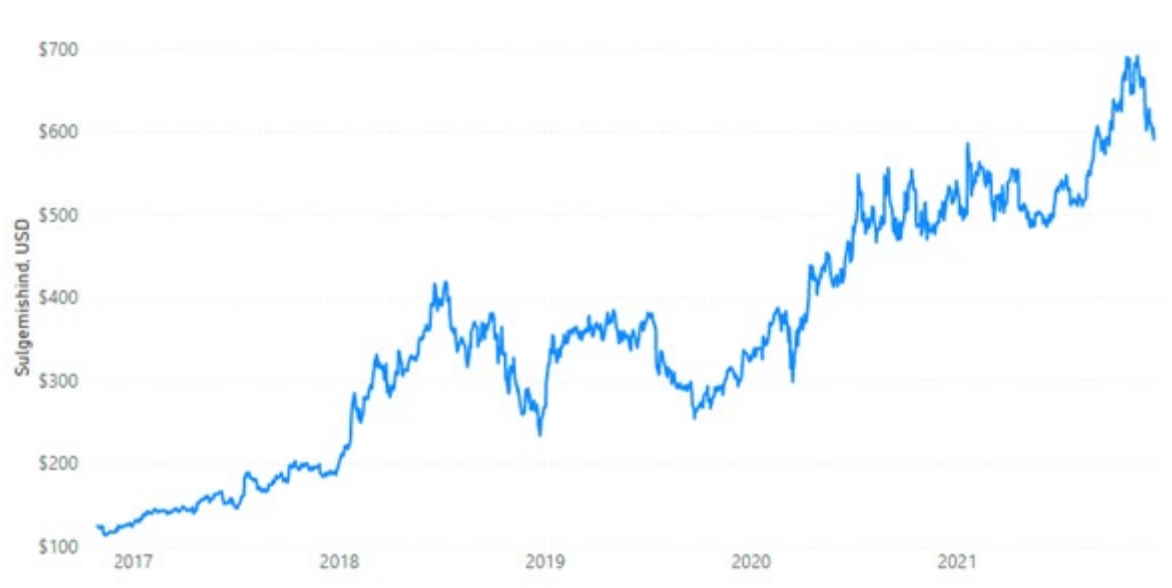
Date	Open	High	Low	Close*	Adj Close**	Volume
Dec 15, 2021	598.18	605.69	584.51	605.04	605.04	2,866,200
Dec 14, 2021	598.71	602.29	588.13	597.99	597.99	2,984,500
Dec 13, 2021	612.00	612.64	599.52	604.56	604.56	2,517,900
Dec 10, 2021	616.78	617.74	605.88	611.66	611.66	2,747,900
Dec 09, 2021	627.58	630.24	610.44	611.00	611.00	2,376,300

Joonis 2.1 Netflix aktsiahindade andmestikust näide

- *Date* – kuupäev.
- *Open* – avanemishind, tähendab hinda, millega aktsiaga kauplemine antud päeval algas.
- *High* – aktsia kõrgeim hind päeva jooksul.
- *Low* – aktsia madalaim hind päeva jooksul.
- *Close* – sulgemishind, tähendab hinda, millega aktsiaga kauplemine antud päeval lõpes.
- *Adj Close* – korporatiivseid tegevusi arvestades korrigeeritud sulgemishind.
- *Volume* – päeva jooksul kaubeldavate aktsiate arv.

Käesolevas töös mudelite treenimiseks kasutatakse Netflix aktsiahindade andmed, mis on võetud ajavahemikul alates 31.10.16 kuni 29.10.21. Esialgselt andmestikust, kasutatakse ainult kahe tulpa, kuupäeva (*Date*) ja sulgemishinna (*Close*).

Testimiseks kasutatakse andmeid, mis on võetud ajavahemikul 01.11.21-16.12.21.



Joonis 2.2 Treenimis- ja testimisandmed: NFLX aktsia sulgemishinnad 31.10.16-16.12.21

Käesoleva lõputöö treenimis- ja testimisandmete aegrida on esitatud Joonisel 2.2.

2.2 Aegridade prognoosimise vahendid POWER BI-s

Käesolevas lõputöös on kirjeldatud aegridade prognoosimise võimalused ja protsess Power BI Desktop tarkvaras.

Prognoosimine Power BI-s põhineb aegridade ennustamise meetodil, mida nimetatakse eksponentsilumiseks. Prognoosid arvutatakse eelmiste väärtuste kaalutud keskmiste abil, kus kaalud vähenevad eksponentsiaalselt ning väikseimad kaalud on seotud rea vanimate väärtustega. [3]

Aastate jooksul on aegridade analüüsimiseks välja töötatud palju meetodeid, olenevalt sellest, kas andmed on tugevalt sesoonsed või puuduvad hooajalisus, kui palju on andmetes müra ja kas andmed sisaldavad "üllatusi" või ebaregulaarseid tippe. Eksponentsilumise meetodil on head tulemused nii teadusuuringutes kui ka ettevõtluses ning selle eeliseks on see, et see summutab müra või soovimatuid variatsioone, mis võivad mudelit moonutada, püüdes samal ajal trende tõhusalt. [10]

Power BI pakub kahte eksponentsiaalse silumise versiooni, ühte sesoonsusega andmete jaoks (ETS AAA) ja teist sesoonsuseta andmete jaoks (ETS AAN). Power BI kasutab automaatselt sobivat mudelit aegrea andmete analüüsi põhjal. [10]

Sesoonsusega algoritm (ETS AAA)

Sesoonsusega algoritm (ETS AAA) modelleerib aegridu, kasutades eksponentsilumise võrrandit, mis arvestab aditiivse vea, aditiivse trendi ja aditiivse sesoonsusega. Seda algoritmi tuntakse mudeli omadusi kirjeldanud teadlaste järgi ka Holt-Wintersi

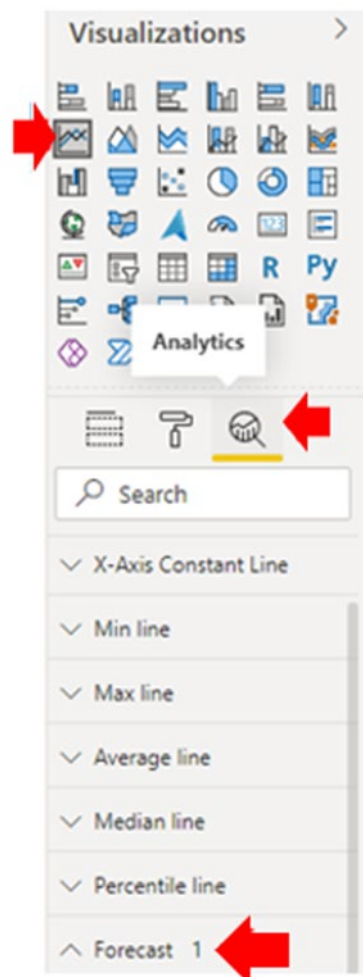
algoritmina. Holt-Wintersi meetodit kasutatakse laialdaselt näiteks ettevõtete nõudluse prognoosimisel ja planeerimisel. [3, 10]

Sesoonsuseta algoritm (ETS AAN)

Sesoonsuseta algoritm (ETS AAN) kasutab aegridade modelleerimiseks lihtsustatud eksponentsilumise võrrandit, mis sisaldab ainult aditiivse trendi ja aditiivse vea komponenti ning ei võta sesoonsust üldse arvesse. Eeldatakse, et aegrea väärtused suurenevad või vähenevad mingil viisil, mida saab kirjeldada valemiga, kuid et suurenemine või vähenemine ei ole tsükliline. [10]

2.2.1 Power BI visuaalid aegridade prognoosimiseks

Vaikimisi Power BI-s prognoosimiseks on ainult aegrea joondiagrammile kuuluv analüütiline vahend, mis on leitav menüüst Analytics → Forecast (vt Joonis 2.3).



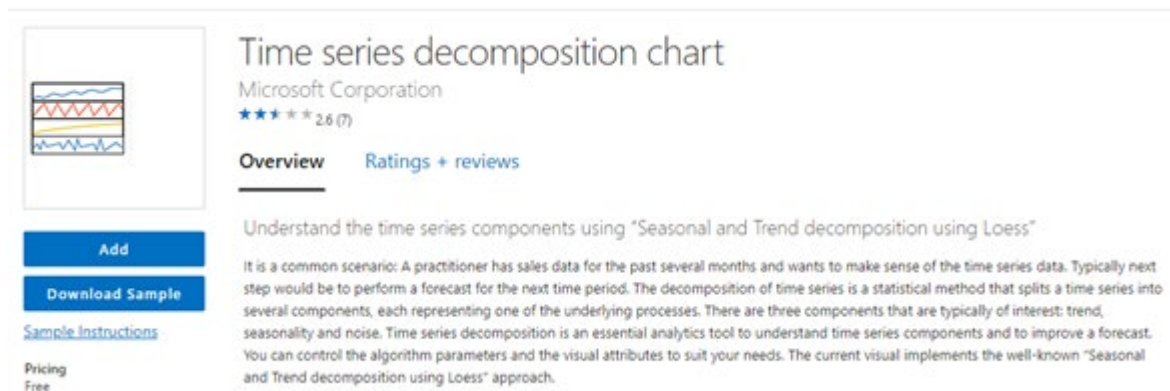
Joonis 2.3 Power BI-sse aegridade prognoosimiseks sisseehitatud vahend

Joondiagramm pakub aegrea prognoosimiseks sesoonsusega või sesoonsuseta algoritmi, mis valitakse automaatselt vastavalt andmetele.

Kuid Power BI võimaldab alla laadida *Microsoft AppSource* ressursilt juurde visuaale, millega saab aegridu prognoosida. Allpool on välja toodud mõned neist.

Visuaal Time Series Decomposition Chart

Time Series Decomposition Chart on avatud lähtekoodiga tasuta Power BI visuaal, mis on väljaantud Microsoft Corporation-i poolt (vt Joonis 2.4). Visuaal on loodud programmeerimiskeele R pakettide *proto* ja *zoo* abil. Praegu saab alla laadida versiooni 1.0.3.0, ning viimati on uuendatud 18.10.2021. [11]



Joonis 2.4 Visuaal *Time Series Decomposition Chart* ressursil *Microsoft AppSource*

Aegridade dekompositsioon on statistiline meetod, mis jagab aegread mitmeks, tavaliselt kolmeks komponendiks: trend, sesoonne komponent ja müra. Aegridade dekompositsioon on oluline analüüsitööriist aegrea komponentide mõistmiseks ja prognoosi parandamiseks. Saadaval on seitse erinevat aegridade visualiseerimise režiimi, mis võimaldavad analüütikul andmete eri aspekte süveneda. Kogunud kasutaja saab kontrollida algoritmi sisemisi parameetreid. [11]

See visuaal aitab paremini mõista aegridu, vaadates hooajalisi mustreid ja trende. Mõistes paremini aegridade keerukust ja aluseks olevaid mustreid, on kergem valida prognoosimudelite parameetreid, neid häälestada ja prognoosi täpsust parandada.

Visuaal Forecasting with ARIMA

Forecasting with ARIMA on avatud lähtekoodiga tasuta Power BI visuaal, mis on väljaantud Microsoft Corporation-i poolt (vt Joonis 2.5). Visuaal on loodud R pakettide *scales*, *forecast*, *zoo*, *htmlWidgets*, *XML* ja *plotly* abil. Praegu saab alla laadida versiooni 1.2.1.0, ning viimati on uuendatud 18.10.2021. [13]

Forecasting with ARIMA
Microsoft Corporation
★★★★☆ 2.8 (21)

Overview Ratings + reviews

Predict future values based on historical data using Autoregressive Integrated Moving Avg (ARIMA)

Use forecasting today to optimize for tomorrow! Time series forecasting is the use of a model to predict future values based on previously observed values.

It is one of the prime tools of any business analyst used to predict demand and inventory, budgeting, sales quotas, marketing campaigns and procurement. Accurate forecasts lead to better decisions. Current visual implements well known Autoregressive Integrated Moving Average (ARIMA) method for the forecasting. ARIMA models are general class of models for forecasting a time series which can be made to be "stationary". While exponential smoothing models are based on a description of trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data. Both seasonal and non-seasonal modeling is supported. You can control the algorithm parameters and the visual attributes to suit your needs.

Add
Download Sample

[Sample Instructions](#)

Pricing: Free
Products: Power BI visuals

Joonis 2.5 Visuaal *Forecasting with ARIMA* ressursil *Microsoft AppSource*

Visuaal rakendab prognoosimiseks ARIMA meetodit, mille eesmärgiks on kirjeldada andmetes esinevaid autokorrelatsioone. Toetatakse nii hooajalist kui ka mittehooajalist modelleerimist. Visuaal võimaldab automaatselt leida hooajalise teguri, kuid lubab ka kasutajale seda määrata. Vaikimisi optimeerib algoritm mudeli kõigi parameetrite jaoks, kuid kasutajal on võimalus muuta mudeli parameetreid. [13]

Visuaal Forecasting using Neural Network

Forecasting using Neural Network on tasuta Power BI visuaal, mis on välja antud MAQ LLC Software poolt (vt Joonis 2.6). Visuaal on loodud R pakettide forecast, plotly, zoo, ja nnet abil. Hetkel on kasutusel 3.0.6.0 versioon, mis on viimati uuendatud 18.10.2021. [14]

Forecast using Neural Network by MAQ Software
MAQ LLC
★★★★☆ 2.9 (11)

Overview Ratings + reviews

Use the Neural Network algorithm to forecast future values based on historical data.

Forecasting using Neural Network by MAQ Software implements an "Artificial Neural Network" to learn from historical data and predict future values. This visual uses a single layer feed forward network with lagged inputs to process time series values. R package dependencies (auto-installed): forecast, plotly, zoo, xts.

Forecasting using Neural Network by MAQ Software is useful for forecasting budgets, sales, demand, or inventory.

Add
Download Sample

[Sample Instructions](#)

Pricing: Free
Key features:

Joonis 2.6 MAQ LLC Software visuaal *Forecasting using Neural Network*

Visuaal kasutab aegrea väärtuste ennustamiseks ühekihilist pärilevivõrku (*feed-forward network*) lubades sisendandmetena antud aegrea suhtes ajas tagasi nihutatud aegridasid (*lags*). Visuaal võimaldab ka mudeli parameetreid käsitsi reguleerida. [14]

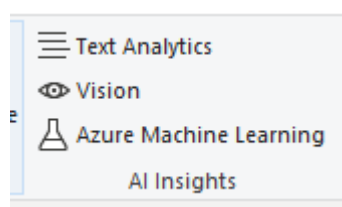
2.2.2 Masinõppe vahendid Power BI-s

Azure masinõppe mudelid Power BI-s

Azure Machine Learning on pilveteenus masinõppeprojekti elutsükli kiirendamiseks ja haldamiseks. Masinõppe spetsialistid, andmeteadlased ja insenerid saavad seda kasutada oma igapäevastes töövoogudes mudelite treenimiseks ja juurutamiseks. *Azure Machine Learning Studio (Azure ML Studio)* on keskkond, kus on võimalik luua masinõppe mudeleid või kasutada avatud lähtekoodiga platvormilt (nt *Pytorch*, *TensorFlow* või *scikit-learn*) loodud mudelit. Keskkonna tööriistad lubavad mudeleid jälgida, ümber õpetada ja ümber paigutada. Azure'i masinõppes on võimalik mudeli treenimiskripti käivitada pilves või luua mudeli nullist. Automatiseeritud masinõppe (*AutoML*) kiirendab seda protsessi ja seda saab kasutada *Azure ML Studio* kasutajaliidese või Pythoni SDK (*Software Development Toolkit*) kaudu. [15]

Antud töös on kasutatud *Azure ML Studio* aegridade prognoosimise automatiseeritud masinõppe algoritmid: *AutoArima*, *Prophet* ja närvivõrgu mudel *ForecastTCN*. Parimaks mudeliks on valitud *AutoArima*, mis on Azure poolt automaatselt häälestatud ARIMA mudel.

Power BI-s on võimalik kasutada Azure ML Studio treenitud ja juurutatud mudelit. Selleks on vaja minna *Power Query Editor*-isse, ning seal valida Azure masinõppe nuppu (vt Joonis 2.7).



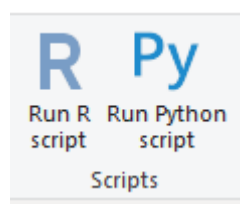
Joonis 2.7 Azure masinõppe nupp Power Query Editor-is

Nupule vajutades, avaneb aken, kus on võimalik näha kõik Azure ML Studios juurutatud mudeleid, ning rakendada neid andmete peale.

Pythoni skriptid Power BI-s

Power BI lubab kasutada programmeerimiskeelte R ja Python skripte andmete sisestamiseks, töötlemiseks ja visualiseerimiseks. R ja Pythoni skriptidega töötamise võimalus avab Power BI-s palju võimalusi ja üks neist on töötada masinõppe tööriistadega ja luua mudeleid otse Power BI-s. Antud lõputöös on kirjeldatud Pythoni keeles skriptide kasutamine aegridade prognoosimiseks.

Pythoni keeles skript masinõppe mudeli loomise koodiga integreeritakse Power Query Editoris andmetabeli andmete teisendamise etapina vajutades Pythoni skriptide käivitamise nupule (vt Joonis 2.8).



Joonis 2.8 Pythoni skripti integreerimise nupp (paremalt)

Pythoni skriptide käivitamisel on mõned piirangud. Neist olulisemad on:

- Imporditakse ainult *pandas* paketi andmetabelid.
- Iga Pythoni skript, mis töötab kauem kui 30 minutit, aegub. [16]

3 TÖÖ TULEMUSED

3.1 Andmete eeltöötlus

Andmete eeltöötlemiseks ja modelleerimiseks kasutatakse Power BI Query Editor-i. Siin jäadi analüüsimisele andmestiku veerud kuupäev (*Date*) ja sulgemishind (*Close*) ning kontrolliti, et kuupäev (*Date*) on kuupäeva formaadis ja sulgemishind (*Close*) on fikseeritud kümnendarvu formaadis.

Selleks, et Power BI-s olevad andmed oleksid värsked, eelkõige on vaja seadistada andmete automaatvärskendamise. Yahoo.Finance portaalist ajaloolised andmed saadakse URL kaudu. Selleks, et ei peaks uute andmete saamiseks URL-i uuesti sisestada, Power Query Editor-is on võimalik seadistada andmestiku automaatvärskendamist.

Yahoo.Finance lehelt kopeeritakse ajalooliste andmete lingi URL aadressiga. Power BI-sse lisades uue allika, valitakse Veeb allikat, ning lisatakse kopeeritud lingi aadressi.

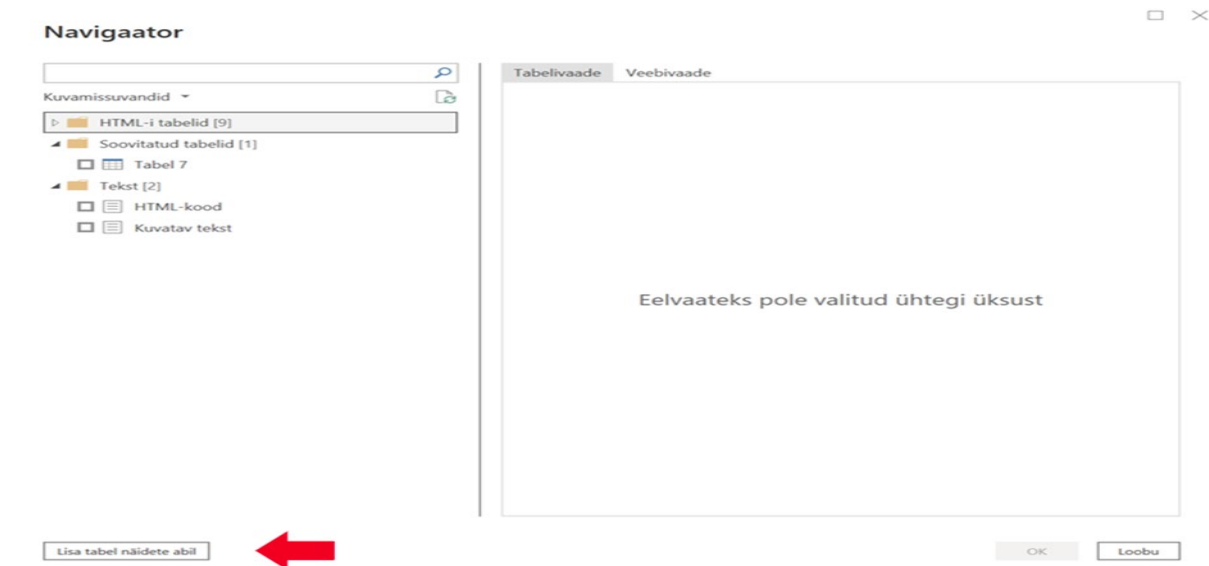
URL aadress näeb välja selliselt:

```
https://query1.finance.yahoo.com/v7/finance/download/NFLX?period1=1604102400&period2=1635465600&interval=1d&events=history&includeAdjustedClose=true
```

Kus NFLX on aktsia kood ja periood on (period1=1604102400&period2=1635465600) epohhi formaadis.

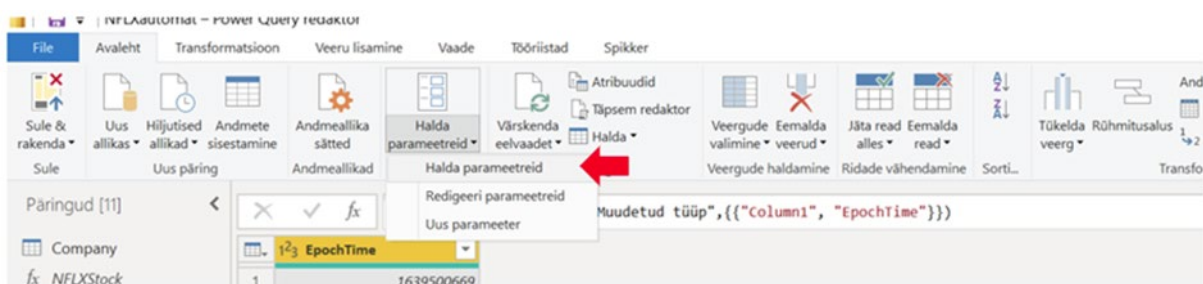
Kuna on vaja, et aruanne oleks alati ajakohane, peaks iga kord, kui klõpsatakse nupul „Värskenda“, asendama URL-i perioodi 2. osa praeguse ajastu ajaga. Selleks, et mitte teha seda käsitsi, lisame andmete juurde tabeli, mis hakkab portaalist <https://www.epochconverter.com/>, võtma viimase epohh aja, ning asendama seda period2 ajaga.

Juurde lisatakse veel üks veebi allikas, aadressiga <https://www.epochconverter.com/>. Ning „Navigaatoris“ lisatakse uus tabel näidete abil (vt Joonis 3.1)



Joonis 3.1 Ajastu aja tabelisse lisamine

Avatud aknasse, Veergu 1 lisatakse kuupäevale vastava ajastu aja numbr. Järgmisena on vaja koostada parameetrit viimase kuupäeva jaoks (period2). On vaja klõpsata päringuredaktoris nuppu „Halda parameetreid“ (vt Joonis 3.2).



Joonis 3.2 Uue parameetri loomine

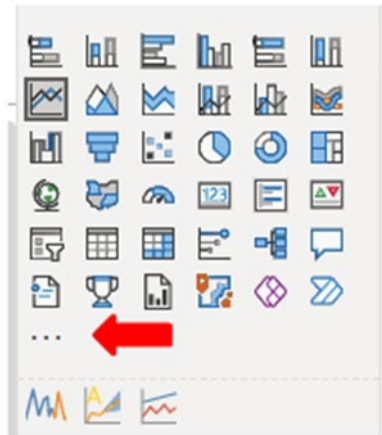
Seal luuakse parameeter *LastDate* ning väärtuseks pannakse ajastu aja numbr (epoch).

Peale seda avatakse andmestiku täpsem redaktor (*Advanced Editor*), kus asendatakse period2 numbriline väärtus parameetri *LastDate* vastu (vt Joonis 3.3).

Tulemusena saadakse tabelisse uus veerg, mis hiljem tuleb laiendada uutele ridadele. Edaspidi andmete värskendamiseks tuleb vajutada nuppu „Värskenda“ Power BI Desktop-is.

3.2 Modelleerimine Power BI visuaalide abil

Selleks, et juurde laadida aegridade prognoosimise visuaalid, visualiseeringute menüüs on vaja vajutada kolme punkti peale, kus valida „Hangi rohkem visuaale“ (vt Joonis 3.6)

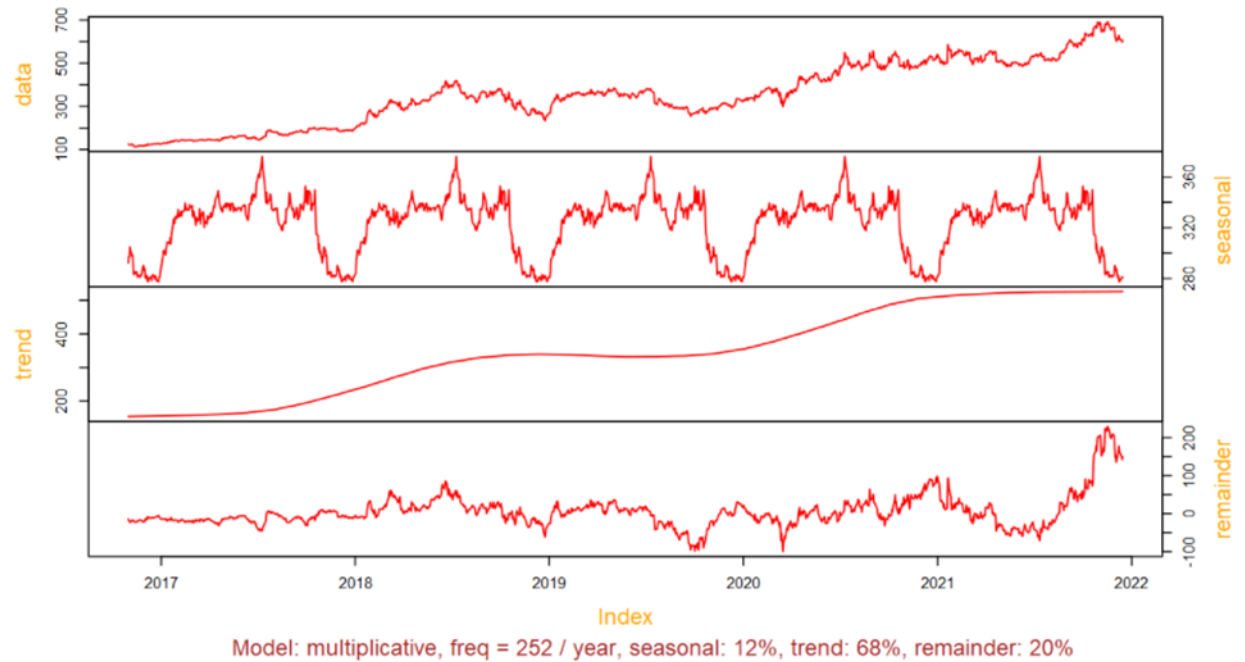


Joonis 3.6 Visuaalide juurde allalaadimine

Pärast seda avaneb Power BI *Microsoft AppSource* ressursi uute visuaalide akna, kus otsingutest saab alla laadida järgmised visuaalid aegridade prognoosimiseks.

Time Series Decomposition Chart

Seda visuaali on vaja enne alla laadida *Microsoft AppSource* kaudu. Joonisel 3.7 on esitatud uuritava aegrea dekomponeerimine.



Joonis 3.7 Netflix aktsia sulgemishindade *Time Series Decomposition Chart*

Aegrea dekomponeerimine tähendab aegrea väärtuste jaotamist järgmisteks mustriteks:

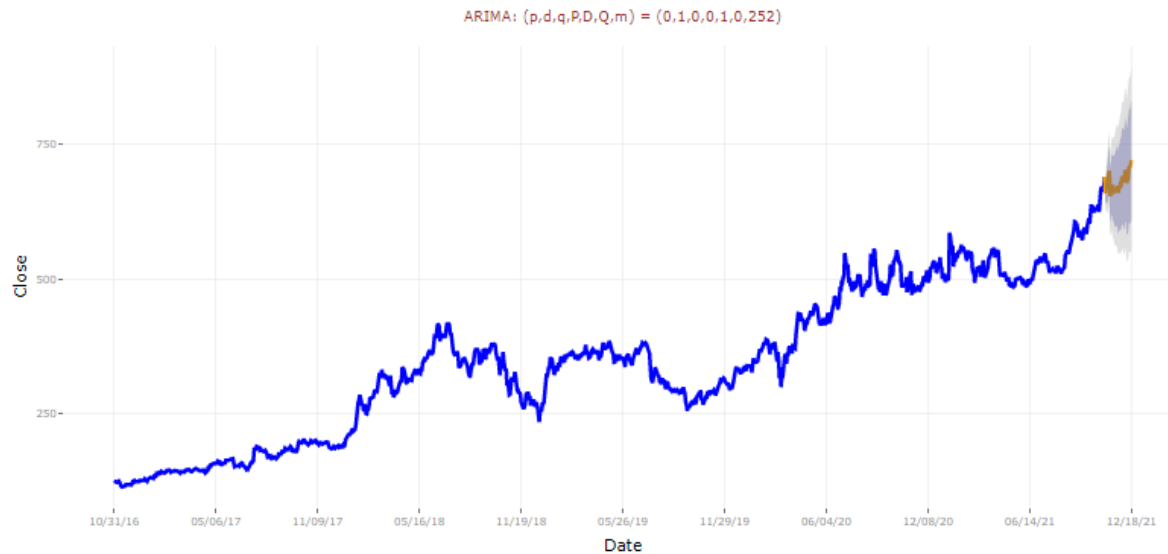
1. Data: esialgne aegrida.
2. Sesoonsus: aegrea perioodiline osa. Uuritavas aegreas esineb sesoone komponent, mis ei muutu aja jooksul, nii et aastatel on sarnased sesoonsed mustrid. Antud andmestiku hooajalisus moodustab 12%.
3. Trend: aegrea väärtustes esinev kasvav või kahanev tendents. Uuritavas aegreas esineb kasvav trend, mis moodustab 68%.
4. Müra ehk jääk: seeria juhuslik komponent moodustab 20% ja näitab viimasel aastal juhuslikkuse kasvava tendentsi.

Sesoonsuse, trendi ja müra kombineerimisel rekonstrueeritakse ülemisel paneelil näidatud esialgsed andmed. Sesoonsuse perioodiks on siin finantsaasta ning ajasammude arvuks on määratud 252, mida saab edasi SARIMA mudelites kasutada.

See visuaal aitab paremini mõista aegridu, vaadates hooajalisi mustreid ja trende.

Forecasting using ARIMA

Seda visuaali on vaja alla laadida *Microsoft AppSource* kaudu.



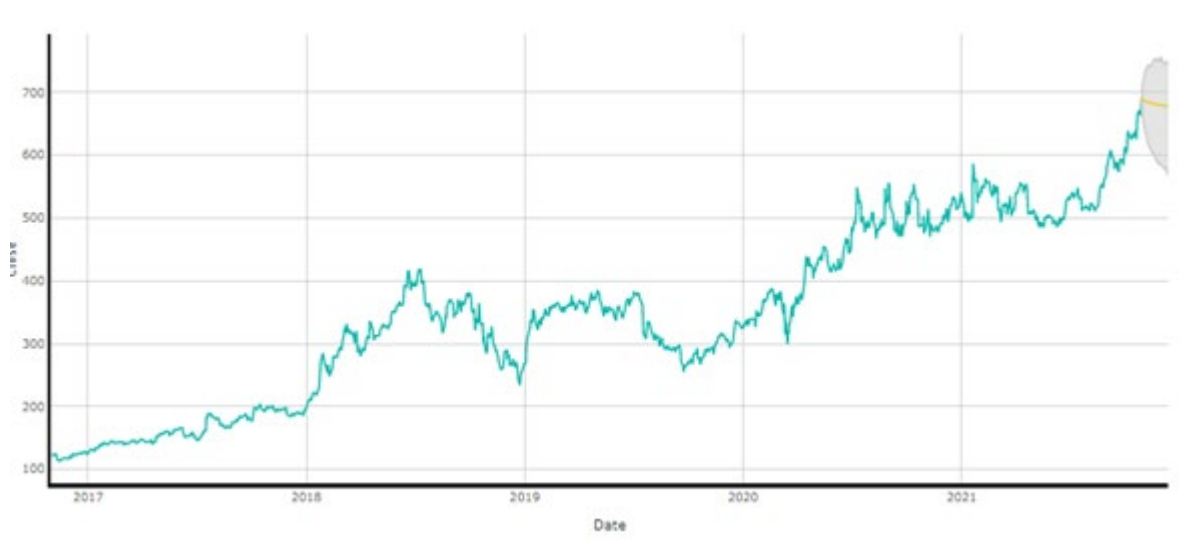
Joonis 3.8 Netflix aktsia sulgemishinna prognoosimine ARIMA(0,1,0)(0,1,0)₂₅₂ mudeliga

ARIMA mudeleid kasutatakse andmestiku autokorrelatsiooni kirjeldamiseks. Autokorrelatsioon on vaatluse korrelatsioon selle varasemate väärtustega. See mõõdab hilinenud vaatluste kogumi sarnasust aja funktsioonina. Joonisel 3.8 on esitatud uuritava aegrea prognoos ARIMA(0,1,0)(0,1,0)₂₅₂ mudeliga, mis näitab aktsiahindade kasvamise tendentsi.

Hallina diagrammil näidatakse prognoosi usaldusvahemik.

Forecasting using Neural Network

Seda visuaali on vaja alla laadida *Microsoft AppSource* kaudu.



Joonis 3.9 Netflix aktsia sulgemishinna prognoosimine närvivõrkuga

Selles visuaalis kasutatakse ühekihilist tehisnärvivõrku, et õppida antud andmete mustreid ja ennustada õpitud parameetrite põhjal tulevasi väärtusi. Joonisel 3.9 on

esitatud uuritava aegrea närvivõrgu mudeliga leitud prognoos, mis erinevalt ARIMA mudelist näitab aktsiahindade kahanemise tendentsi.

Power BI-sse sisseehitatud visuaalide puuduseks on see, et prognooside väärtused ei ole võimalik alla laadida või tabelina salvestada ning nende järgi mudelite RMSE vead leida.

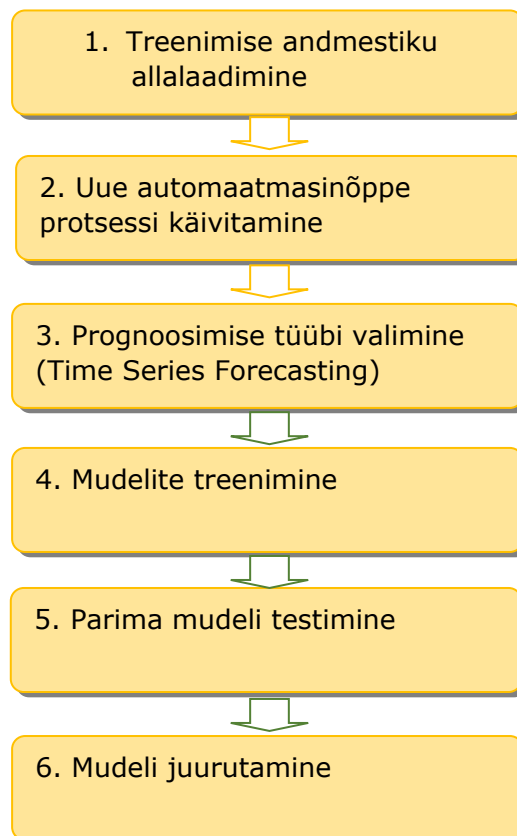
3.3 Automaatmodelleerimine Azure ML Studios

Azure ML Studios prognoosimudeli loomiseks ja juurutamiseks, et neid hiljem saaks kasutada Power BI-s visualiseerimiseks, on valitud mudeli treenimine ja juurutamine automaatse masinõppega. See valik võimaldab täielikult automatiseeritud andmete ettevalmistamist ja mudeli treenimist ilma koodi kasutamata.

Automatiseeritud masinõppe võimaldab optimaalse algoritmi ja parameetrite leidmiseks automaatselt treenida palju erinevaid masinõppe mudeleid.

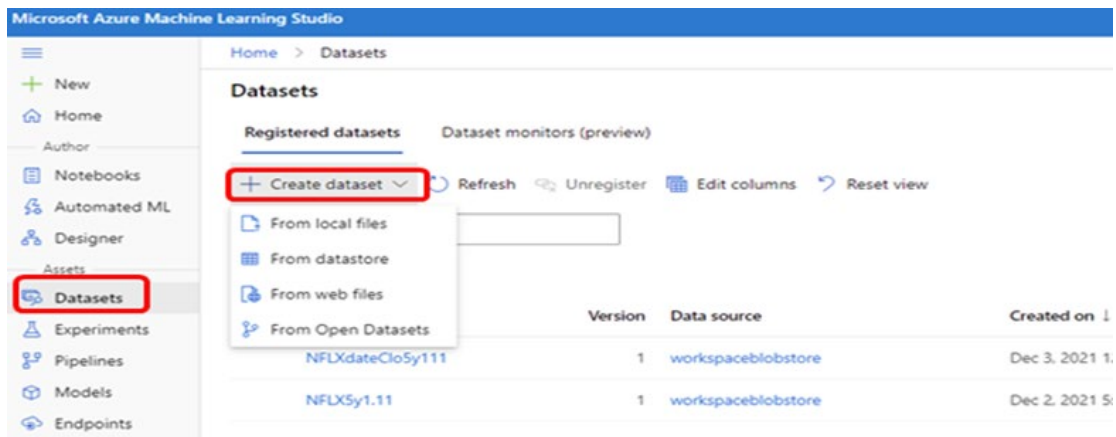
3.3.1 Modelleerimise protsess

Automaatmasinõppega modelleerimise protsess on esitatud Joonisel 3.10.



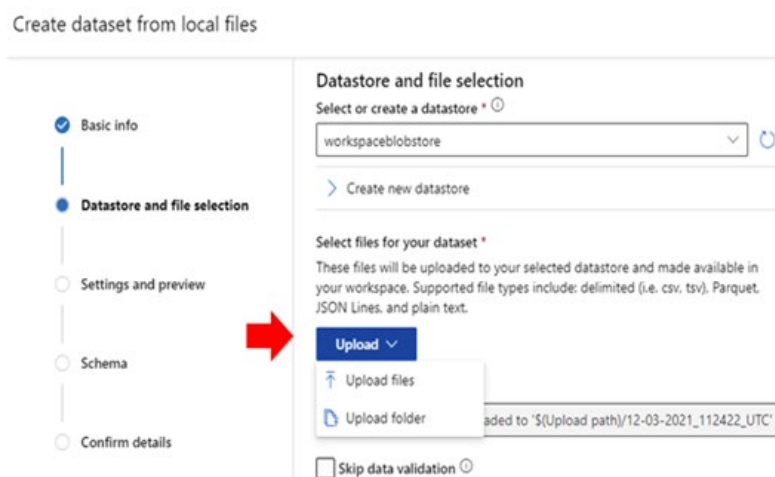
Joonis 3.10 Automaatmasinõppega modelleerimise protsess

Esimese sammuna on vaja laadida treenimisandmestik (vt Joonis 3.11).



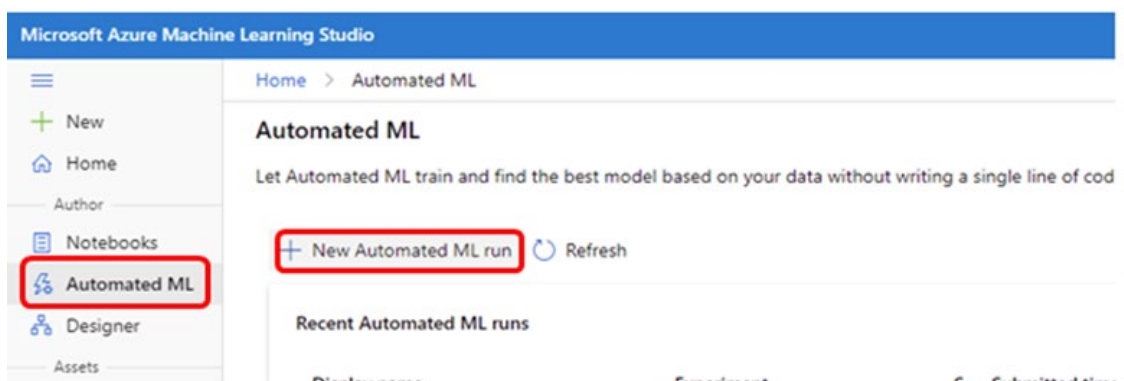
Joonis 3.11 Uue andmestiku allalaadimine

Upload nupule vajutades, tuleb valida *upload files*, et saaks treeningandmetega csv-formaadis faili arvutist Azure ML Studiosse alla laadida (vt Joonis 3.12).



Joonis 3.12 Andmestiku allalaadimine arvutist

Teiseks, Azure ML automaatmasinõppe loomiseks, peab valima vasakul tulbas Automated ML, ning uus automaatne mesinõppe (vt Joonis 3.13)

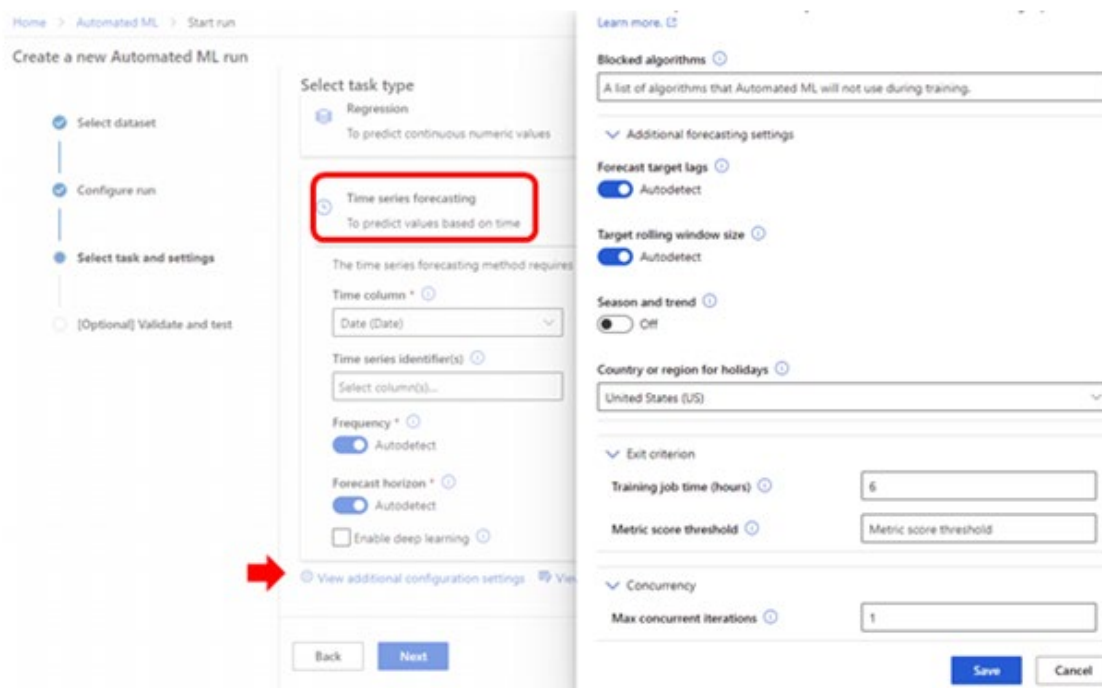


Joonis 3.13 Uue Azure automaatmasinõppe loomine

Kolmas samm on prognoosimise tüübi valimine (vt Joonis 3.14). Azure pakub prognoosimise tüübiks klassifitseerimise, regressiooni või aegridade prognoosimise meetodit. Antud töös kasutatakse aegridade prognoosimise meetodit (*Time Series Forecasting*). *Time Series Forecasting* (Aegridade prognoosimine) - väärtuste prognoosimine aja parameetri järgi. Vajutades all nurgas *Lisaparametrite* peale, on võimalik lisaparametreid konfigurida.

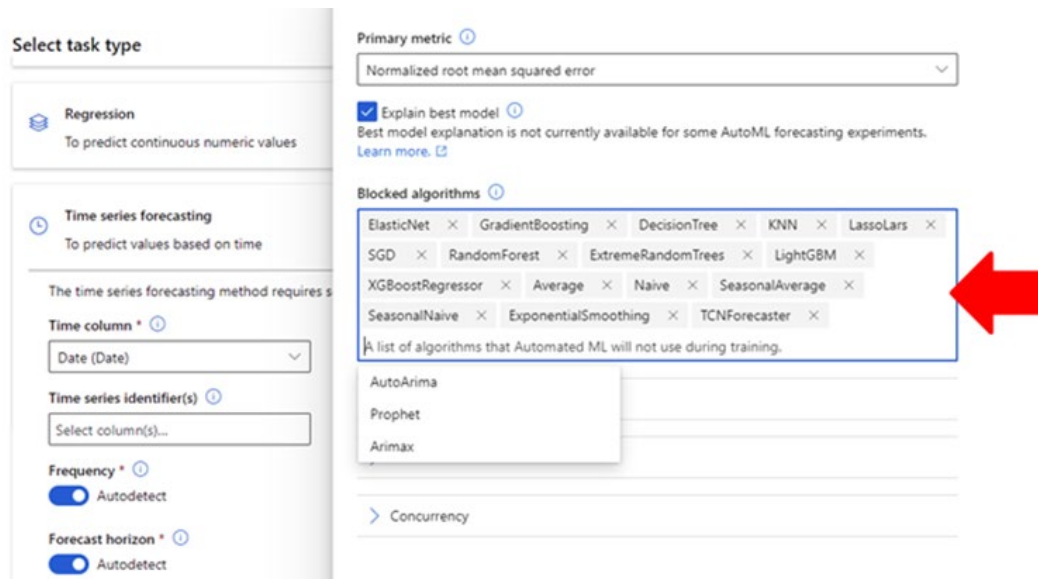
Prognoosimise tegemiseks on vaja

- Lubada sügavõpe.
- Valida aja veerg, ehk *Date*
- Valida prognoosihorisont. See näitab, mitu ajaühikut (minutid, tunnid, päevad, nädalad, kuud või aastad) suudab mudel tulevikku ennustada. Vaikimisi prognoosihorisondiks määratakse 7 päeva.



Joonis 3.14 Prognoosimise tüübi valimine

Neljandaks, kõikide mudelite treenimine võtab aega umbes kaks tundi. Et seda aega vähendada, on võimalik treenimiseks jätta ainult andmetele sobivaid aegridade prognoosimise algoritme, mis on AutoArima, Arimax, Prophet ja ForecastTCN (Azure ML Studio poolt pakutav närvivõrgu mudel). Selleks on võimalik ülejäänud algoritmid välja jätta (vt Joonis 3.15).



Joonis 3.15 Algoritmid blokeerimine

Automaatmasinõppe protsessi lõppemisel pakutakse andmetele sobitud mudelite nimekirja (vt Joonis 3.16).

keen_glove_r36zwck5

Refresh Cancel Delete

Details Data guardrails **Models** Outputs + logs Child runs Snapshot

Refresh Deploy Download Explain model Edit columns Reset view

Search Submitted time All filters Clear all

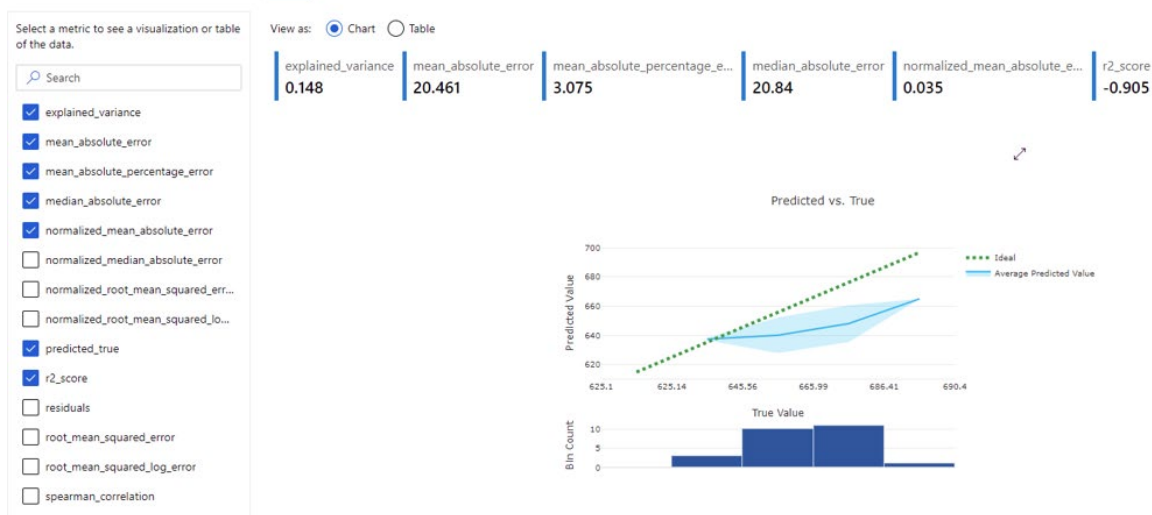
Showing 1-2 of 2 models Page size: 25

Algorithm name	Explained	Normalized ro... ↑	Sampling	Subn
AutoArima	View explanation	0.03827	100.00 %	Dec t
ProphetModel		0.17910	100.00 %	Dec t

Joonis 3.16 Mudelite nimekiri

Neljast algoritmist *Azure Automated ML* konstrueeris ainult kahel algoritmil põhinevad mudelid *ProphetModel* ja *AutoArima*. Parimaks oli pakutud *AutoArima* mudel. Joonisel 3.17 on näha mudeli parameetreid. Treeningandmetel mudeli RMSE on 22.077.

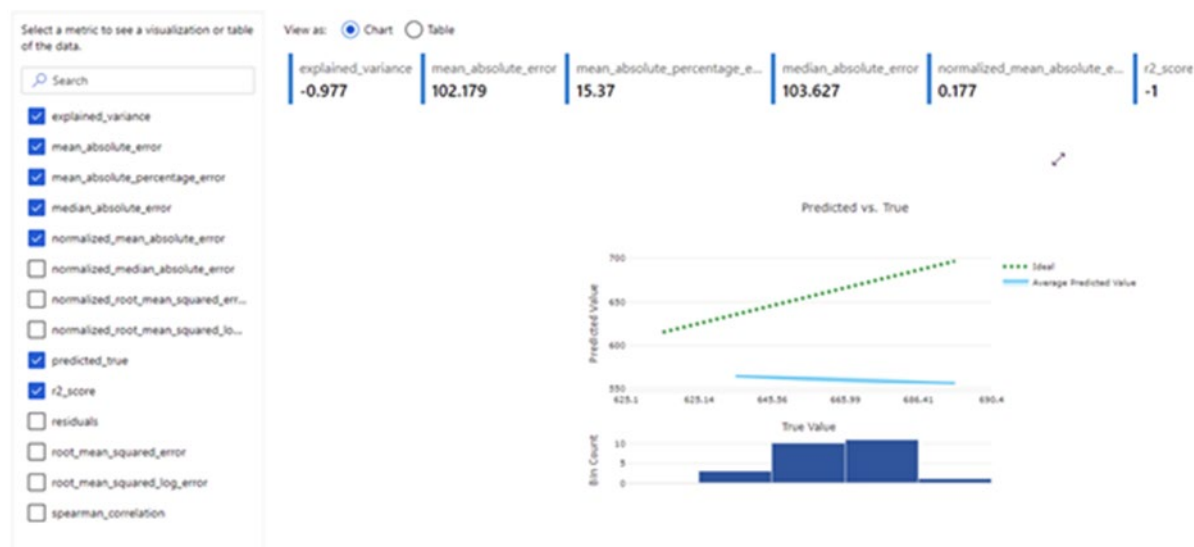
explained_variance	mean_absolute_error	mean_absolute_percentage_e...	median_absolute_error	normalized_mean_absolute_e...	r2_score
0.148	20.461	3.075	20.84	0.035	-0.905
root_mean_squared_error	root_mean_squared_log_error	spearman_correlation			
22.077	0.034	0.76			



Joonis 3.17 AutoArima treeningandmetel mudeli täpsuse mõõdikud

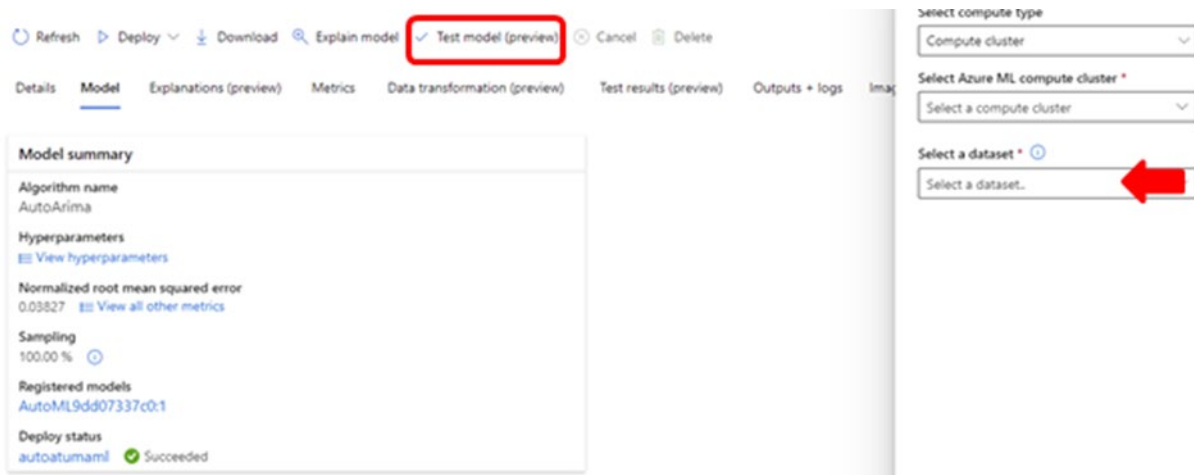
Prophet mudeli RMSE oli 103.627 (vt Joonis 3.18).

explained_variance	mean_absolute_error	mean_absolute_percentage_e...	median_absolute_error	normalized_mean_absolute_e...	root_mean_squared_error	root_mean_squared_log_error
-0.977	102.179	15.37	103.627	0.177	103.33	0.169
normalized_root_mean_squar...	normalized_root_mean_squar...					
0.094	0.179					



Joonis 3.18 Prophet treeningandmete mudeli täpsuse mõõdikud

Viidendaks, enne mudeli juurutamist ja Power BI-sse kasutusele võtmist, Azure ML Studio pakub mudeli testimise võimalust. Testimiseks valitakse testandmestiku (vt Joonis 3.19).



Joonis 3.19 Mudeli testimine testandmetega

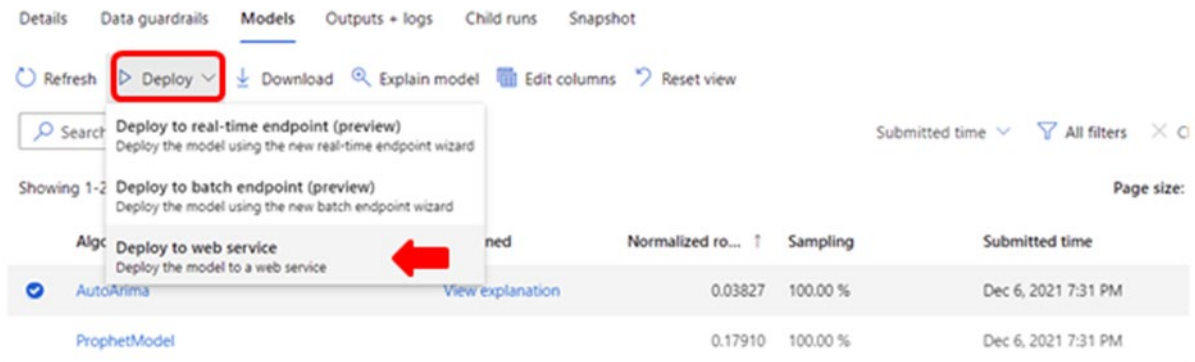
AutoArima mudeli testandmetel testimise tulemused on esitatud Joonisel 3.20. RMSE väärtuseks on juba 44.45.



Joonis 3.20 AutoArima testandmete mudeli mõõdikud

Viimane, kuues samm on parima mudeli juurutamine Power BI-s kasutamiseks. Kui automatiseeritud masinõppe töö on lõpule viidud, saab vaadata kõigi kasutatud masinõppe mudelite loendit, valides vahekaardi Mudelid. Mudelid järjestatakse täpsuse

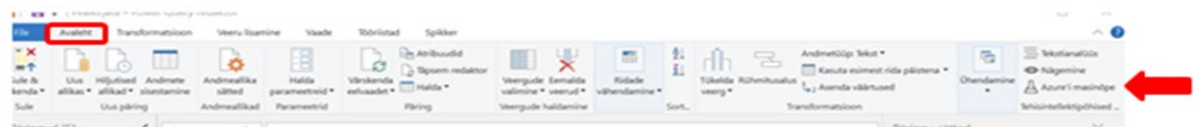
järgi, kõige paremini toimiv mudel, siin *AutoArima*, kuvatakse esimesena. Pärast parima mudeli valimist aktiveeritakse nupp Juuruta (Deploy) (vt Joonis 3.21).



Joonis 3.21 Mudeli juurutamine

3.3.2 Azure mudelite rakendamine Power BI-s

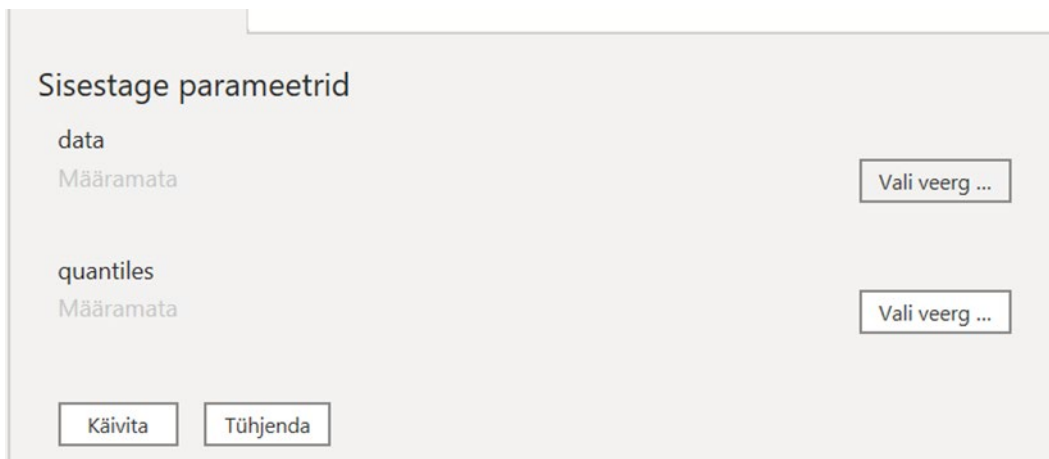
Power BI Query Editor-is on olemas nupp Azure Machine Learning (vt Joonis 3.22).



Joonis 3.22 Azure masinõppe mudelite kasutamine Power BI-s

Enne Azure masinõppe mudeli rakendamist, on vaja teha *Enter Data* abil lisatabeli kvantiili näitudega. Selleks oli lisatud *Quantiles* nimega ühe veeruga ja ühe reaga tabel, kuhu kvantiili väärtuseks oli sisestatud 0.975.

Prognoosi loomiseks, kasutatakse testimisandmestikku. Valitakse Azure ML mudel, parameetriks *data* valikakse testimisandmestikust veerg kuupäev (*Date*) ja parameetriks *quantiles* valitakse kvantiil (vt Joonis 3.23).



Joonis 3.23 Azure ML mudeli parameetrite valimine

Tabelisse lisatakse veerg prognoosidega, kuid prognoosid ei ole numbriliste väärtustega, nad on json -formaadis ning neid on vaja arvulise formaadi konverteerida. Selleks on vaja lisada uue kohandatud veergu, ning valemiks kasutada `Record.ToTable(Json.Document(["AzureML.autoatumaml"]))`. Lisandub uus veerg *Custom*.

Seejärel tuleb lisada peatabelisse uue indeksite veergu ning teha tabelist duplikaati. Duplikaati saab teha vajutades parema hiireklõpsuga tabeli nime peale. Edasised toimingud tehakse dubleeritud tabelis.

Uues veerus *Custom* klõpsata esimese rea peale (vt Joonis 3.24)

ded Custom", "Index", 1, 1, Int64.Type)		123 Custom	123 Index
0586836, 329.8578061273216, 329.3642516...	Table		1
0586836, 329.8578061273216, 329.3642516...	Table		2
0586836, 329.8578061273216, 329.3642516...	Table		3
0586836, 329.8578061273216, 329.3642516...	Table		4
0586836, 329.8578061273216, 329.3642516...	Table		5
0586836, 329.8578061273216, 329.3642516...	Table		6
0586836, 329.8578061273216, 329.3642516...	Table		7
0586836, 329.8578061273216, 329.3642516...	Table		8
0586836, 329.8578061273216, 329.3642516...	Table		9
0586836, 329.8578061273216, 329.3642516...	Table		10
0586836, 329.8578061273216, 329.3642516...	Table		11
0586836, 329.8578061273216, 329.3642516...	Table		12

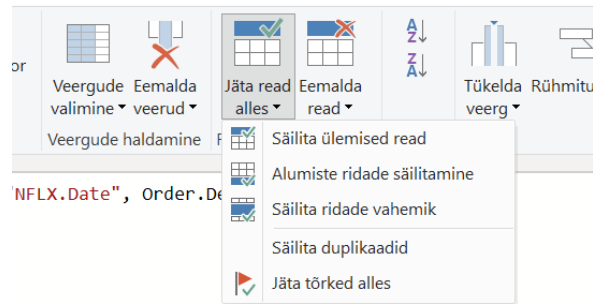
Joonis 3.24 Kohandatud tabelis prognooside avamine

Avaneb tabel prognoosidega, millest tuleb alles jätta ainult esimene rida (vt Joonis 3.25).

Name	Value
1 forecast_interval	List
2 prediction_interval	List
3 index	List

Joonis 3.25 Tabel prognoosidega

Et jätta ainult esimest rida, võib kasutada *Jätta read alles* ikooni, ning valida säilita ülemised read (vt Joonis 3.26)



Joonis 3.26 Ülemiste ridade säilitamine

Edasi tuleb teha lahti loetelu. Lisada tabelisse indeksiveerg, ning ühendada duplikaat tabeliga (*Combine -> Merge Queries*), jättes lõpptabelisse ainult prognoosid ning kuupäeva (*Date*) ja sulgemishinna (*Close*) veerud.

Tulemusena on võimalik Power BI-s saada visualiseeringu (vt Joonis 3.27), kus helesinine joon on prognoositav sulgemishind.



Joonis 3.27 AzureML AutoArima mudeli prognoos ja tegelikud väärtused

Prognoosid ja tegelikud väärtused on võimalik Power BI-s ka tabelina esitada (vt Joonis 3.28).

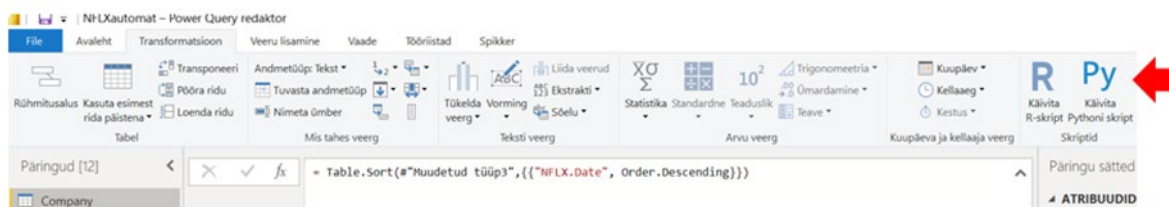
Date	Forecast	Close
1. november 2021. a.	\$687,28	\$681,17
2. november 2021. a.	\$687,55	\$677,72
3. november 2021. a.	\$688,63	\$688,29
4. november 2021. a.	\$688,40	\$668,40
5. november 2021. a.	\$688,92	\$645,72
8. november 2021. a.	\$688,46	\$651,45
9. november 2021. a.	\$690,58	\$655,99
10. november 2021. a.	\$691,46	\$646,91
11. november 2021. a.	\$689,91	\$657,58
12. november 2021. a.	\$690,60	\$682,61
15. november 2021. a.	\$691,07	\$679,33
16. november 2021. a.	\$692,86	\$687,40
17. november 2021. a.	\$692,54	\$691,69
18. november 2021. a.	\$691,55	\$682,02
19. november 2021. a.	\$693,04	\$678,80
22. november 2021. a.	\$693,24	\$659,20
23. november 2021. a.	\$694,72	\$654,06

Joonis 3.28 AutoArima prognoosiandmed ja tegelikud väärtused tabelina

3.4 Modelleerimine Pythoni abil

3.4.1 Pythoni skriptid Power BI-s

Pythoni skriptide kasutamiseks Power BI-s, peab olema Python ja selle kõik vajalikud teigid arvutisse installitud, ning Power BI seadistuses (*Options*) Pythoni ja Pythoni integreeritud programmeerimiskeskonna (*IDE, Integrated Development Environment*) asukoht arvutis määratud. Power Query editoris on olemas *Run Python script* nupp, kuhu peale vajutades avaneb aken, kuhu on võimalik Pythoni skripti sisestada (vt Joonis 3.29). [16]

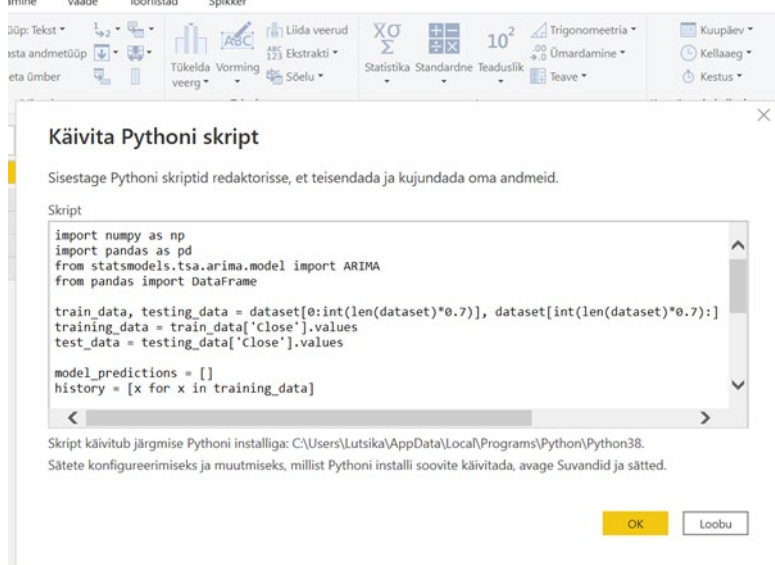


Joonis 3.29 Pythoni skripti käivitamine

3.4.2 Pythoni mudelite rakendamine Power BI-s

Enne Pythoni skripti käivitamist, tuleb andmestiku modelleerida. Esialgsest andmestikust tehakse viide, kasutatavad veerud *Date* ja *Close* asendatakse numbrilise

tüübi vastu. Seejärel avaneb aken, kuhu saab lisada Pythoni ARIMA mudeli skript (vt Joonis 3.30).



Joonis 3.30 Pythoni skripti lisamine

Kui Pythoni skript on läbi töödeldud, ilmub ette uus tabel skripti väljundandmetega (vt Joonis 3.31), kus *dataset* on koguandmestik, *df_c* prognoosid, *testing_data* on testimisandmestik ja *train_data* on treeningandmestik.

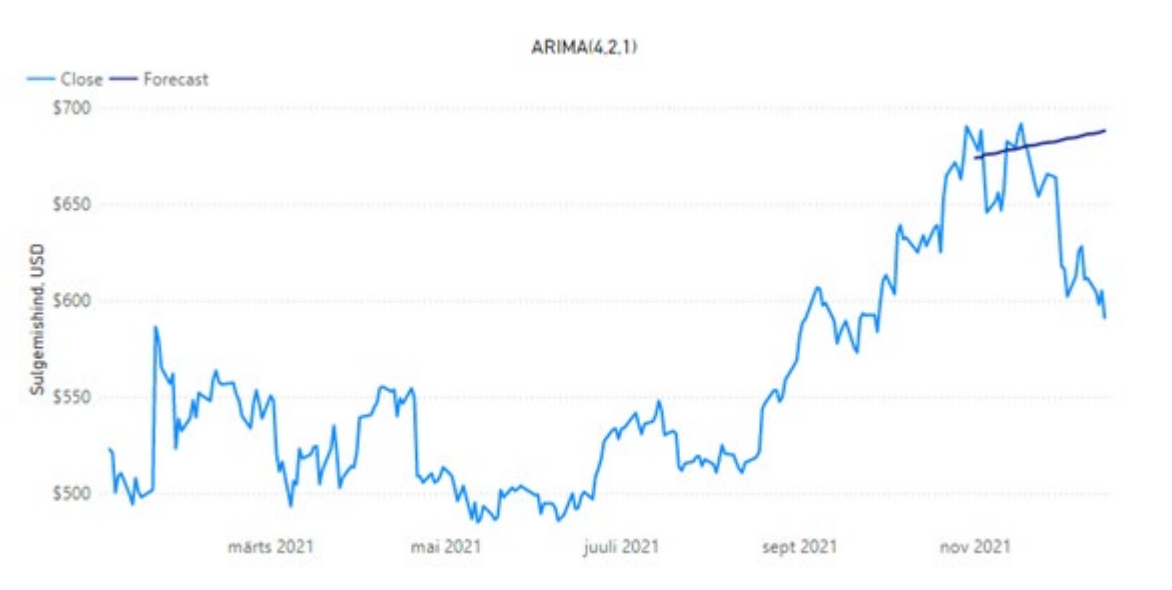
	Name	Value
1	dataset	Table
2	df_c	Table
3	testing_data	Table
4	train_data	Table

Joonis 3.31 Pythoni ARIMA mudeli rakendamise andmestik

Vajutades *df_c* Table peale avaneb tabel koos prognoosimise andmetega. Nimega 0 veerg ümber nimetatakse prognoosi veeruks (*Forecast*)

Avatud tabelis tuleb muuta veergu väärtuste tüübid. *Date* pannakse kuupäeva formaati, *Close* ja *Forecast* veergudele määratakse fikseeritud kümnendarvu tüübi.

Tulemusena saadud Pythoni skripti ARIMA mudeli prognoosid ja aegrea tegelikud väärtused on esitatud Joonisel 3.32.



Joonis 3.32 Pythoni skripti ARIMA mudeli prognoos ja tegelikud väärtused

Pythoni abil leitud prognoosid ja tegelikud väärtused on võimalik Power BI-s ka tabelina esitada (vt Joonis 3.33).

Date	Close	Forecast
1. november 2021. a.	\$681,17	\$673,98
2. november 2021. a.	\$677,72	\$674,26
3. november 2021. a.	\$688,29	\$674,09
4. november 2021. a.	\$668,40	\$675,55
5. november 2021. a.	\$645,72	\$675,83
8. november 2021. a.	\$651,45	\$676,28
9. november 2021. a.	\$655,99	\$676,65
10. november 2021. a.	\$646,91	\$677,19
11. november 2021. a.	\$657,58	\$677,60
12. november 2021. a.	\$682,61	\$678,04
15. november 2021. a.	\$679,33	\$678,47
16. november 2021. a.	\$687,40	\$678,92
17. november 2021. a.	\$691,69	\$679,35
18. november 2021. a.	\$682,02	\$679,79
19. november 2021. a.	\$678,80	\$680,22
22. november 2021. a.	\$659,20	\$680,66
23. november 2021. a.	\$654,06	\$681,10

Joonis 3.33 Pythoni ARIMA mudeli prognoosiandmed ja tegelikud väärtused tabelina

Saadud tabeli on võimalik edasi csv-formaadis eksportida või otse Power BI-s moodsiku abil testandmete RMSE arvutada. Pythoni ARIMA mudeli testandmete RMSE on 48.47, mis on Azure AutoArima mudeliga lähedane. Pythoni ARIMA mudeli parameetreid saab edasi hallata, skripti uuesti käivitada ning proovida muuta mudel täpsemaks.

KOKKUVÕTE

Üha rohkem ettevõtted võtavad kasutusele erinevaid *Business Intelligence* (BI) tarkvarasid, mille abil on võimalik andmeid visualiseerida ja analüüsida ning luua masinõppe abil prognoosimise mudeleid. Käesoleva lõputöö põhieesmärgiks oli uurida Power BI aegridade prognoosimise ja masinõppe vahendeid aktsiahindade modelleerimise näitel.

Töös olid uuritud aegridade prognoosimisel kasutatavad mudelid, mis olid loodud Azure Machine Learning Studio automaatmasinõppe abil, kui ka Pythoni skriptide abil ning rakendatud Power BI Desktop-is.

Lõputöö põhiosa koosneb kolmest peatükist. Esimeses peatükis tutvustati aegridade prognoosimise meetodeid ja mudelid. Lisaks esitati kasutatud valideerimise moodsid.

Teises peatükis anti ülevaade töös kasutatud andmete kohta ning Power BI tarkvara aegridade prognoosimise võimalustest. Tutvustati Power BI-sse sisseehitatud prognoosimise võimalusi. Näidati mõned populaarsemad visuaalid aegridade prognoosimiseks, mida saab juba praegu Power BI-sse alla laadida ja seal kasutada. Tehti ka Power BI masinõppe vahendite ülevaade.

Kolmandas peatükis modelleeriti aktsiahinnad Power BI aegridade prognoosimise meetodite abil, kasutades nii Azure masinõppe kui ka Pythoni skriptide integreerimist ning tehti leitud mudelite visualisatsioonid Power BI-s. Lõputöös esitatud aegridade modelleerimise protsessi kirjeldus on kasutatav Power BI masinõppe võimalustega tutvumiseks Power BI tarkvara täiendkoolituste läbiviimisel.

Pythoni skriptidega töötamise võimalus avab Power BI-s palju võimalusi ja üks neist võimalustest on töötada masinõppe tööriistadega ja luua tasuta erinevat tüüpe mudeleid otse Power BI-s ning koostöömismisel Power BI andmete importimise ja uuendamise võimalustega lubab luua ja visualiseerida mudeleid peaaegu reaalajas.

Aktsiate hindasid on väga raske prognoosida, sest aktsiaturg on väga ettearvamatu, igasugune geopoliitiline muutus võib mõjutada aktsiate dünaamikat börsil. Andsid ka, töökäigus loodud aegridade prognoosimise mudelid umbkaudsed tulemused. Kuid antud mudelitega on võimalik edasi töötada, selleks on igas mudelis ja visuaalis võimalik parameetreid hallata, ning muuta mudel täpsemaks.

SUMMARY

The choice of the topic is due to the author's interest in the stock markets. Stock price modeling and forecasting using Power BI software were chosen because the author has already been in contact with this software and wanted to further develop his knowledge.

This thesis on the topic „Time Series Forecasting on Stock Prices Using Power BI Software“ is written by student Lutsika Klim.

The main goal of this thesis is to study Power BI machine learning tools on the example of stock price modeling. The thesis looks at different time series predictions and further behavior modeling methods and machine learning models offered by Power BI software tools. The accuracy of these models is benchmarked using stock price data and also visualizations of stock prices and their forecasts are provided. Based on the aim of the thesis, the tasks of the thesis have also been set: to study the machine learning methods built into Power BI, which allow time series to be predicted; use them to construct models for modeling stock prices based on data from the Yahoo.Finance portal; compare the accuracy of the obtained models and visualize stock price forecasts based on the best models.

The work consists of three chapters.

- The first chapter introduces the methods for forecasting time series and presents the validation metrics used.
- The second chapter describes the data used in the work and the possibilities of forecasting the time series using Power BI software.
- In the third chapter, stock prices are modeled using Power BI time series prediction methods such as Azure machine learning integration and Python scripts, and visualizations of the best models are made.

The description of the time series modeling process presented in the can be used for educational purposes in TalTech Virumaa College.

The ability to work with Python scripts opens up many possibilities in Power BI, and one of those options is to work with machine learning tools and create different types of models directly in Power BI, which in combination with Power BI data import and refresh capabilities allows to create and visualize models in near real time.

Stock prices are very difficult to predict because the stock market is very unpredictable, any geopolitical change can affect stock market dynamics. The time series prediction models created in the workflow also gave approximate results.

However, it is possible to work with these models, for which it is possible to manage parameters in each model and visual and to make the model more accurate.

KASUTATUD KIRJANDUSE LOETELU

1. Yahoo.Finance. [Online] <https://finance.yahoo.com/> (01.11.2021).
2. History of The Stock Market. [Online] <https://bebusinessed.com/history/history-of-the-stock-market/> (01.11.2021).
3. Hyndman, R.J., Athanasopoulos, G. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia, 2018. [E-raamat] <https://otexts.com/fpp2> (17.12.2021).
4. Time Series Forecasting using Microsoft Power BI. [Online] <https://www.analyticsvidhya.com/blog/2020/07/time-series-forecasting-using-microsoft-power-bi/> (17.12.2021).
5. A Guide to Time Series Forecasting in R You Should Know. [Online] <https://www.simplilearn.com/tutorials/data-science-tutorial/time-series-forecasting-in-r> (17.12.2021).
6. ARIMA & SARIMA: Real-World Time Series Forecasting (Advanced Guide) [Online] <https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide> (17.12.2021).
7. Prophet. [Online] <https://facebook.github.io/prophet/> (17.12.2021).
8. Forecasting: Principles & Practice. [Online] <https://otexts.com/fpp3/prophet.html> (17.12.2021).
9. LSTM Framework For Univariate Time-Series Prediction. [Online] <https://towardsdatascience.com/lstm-framework-for-univariate-time-series-prediction-d9e7252699e> (17.12.2021).
10. Describing the forecasting models in Power View. [Online] <https://powerbi.microsoft.com/en-us/blog/describing-the-forecasting-models-in-power-view/> (17.12.2021).
11. Time Series Decoposition Chart. [Online] <https://appsource.microsoft.com/en-us/product/power-bi-visuals/WA104380897?tab=Overview> (17.12.2021).
12. Time Series Forecasting in PowerBI. [Online] <https://databi.me/2020/03/14/time-series-forecasting-in-powerbi/> (17.12.2021).
13. Forecasting with Arima. [Online] <https://appsource.microsoft.com/en-us/product/power-bi-visuals/WA104380888?tab=Overview> (17.12.2021).
14. Forecast using Neural Network. [Online] <https://appsource.microsoft.com/en-us/product/power-bi-visuals/WA104381428?tab=Overview> (17.12.2021).
15. What is Azure Machine Learning? <https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-machine-learning> (17.12.2021).
16. Run Python scripts in Power BI Desktop. [Online] <https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-python-scripts> (17.12.2021).