

МАТЕМАТИЧКА ГИМНАЗИЈА

МАТУРСКИ РАД

из информатике и програмирања

Предвиђање цене акција на берзи коришћењем
неуронских мрежа

Ученик

Момчило Мркајић
одељење 4д

Ментор

Миодраг Радојевић

Београд, мај 2021.

Апстракт

Предвиђање будућег кретања цена деоница је предмет истраживања већ много година. Са једне стране, имамо људе који верују у хипотезу ефикасног тржишта, чија је директна последица да је немогуће предвидети вредност акција, док са друге стране, постоје предлози да ју је могуће предвидети са високим степеном прецизности. У овом раду предлажемо приступ са машинским учењем. Предвиђана је цена деоница компаније Apple Inc. при затварању трансакција тог дана. Поређена су два модела: CNN и LSTM, при чему јасно видимо да се CNN модел брже прилагођава на промене у маркету.

Abstract

Prediction of future movement of stock prices has been a subject of research for many years. On the one hand, there are people who believe in the Efficient Market Hypothesis, whose direct implication is that it is impossible to predict a value of a stock, while on the other hand, there are propositions that it can be predicted with a high level of accuracy. In this paper, we propose an approach with machine learning. We are predicting the stock price of company Apple Inc. at the end of the fiscal day. We are comparing two models: CNN and LSTM, where we clearly see that CNN model can adapt faster to changes in the market.

Садржај

1	Увод	4
2	Хипотеза	5
3	Метод	5
3.1	RNN	5
3.2	LSTM	6
3.3	CNN	8
3.3.1	Конволуциони слојеви	9
3.3.2	Pooling слој	10
3.3.3	Потпуно повезан слој	10
3.4	Активационе функције	10
3.4.1	Сигмоид	11
3.4.2	Tanh	11
3.4.3	ReLU	11
3.5	База података	12
3.6	Модели	12
4	Метрика	13
5	Резултати	13
6	Закључак	15

1 Увод

Предвиђање се може дефинисати као покушај да се погоди неки догађај у будућности коришћењем историјских података. У многим областима је ово битно, па и у бизнису, економији и финансијама. Може се поделити у кракторочно (предвиђање за наредних неколико секунди па све до неколико месеци) и дугорочно (предвиђање за наредних неколико година). Већина проблема предвиђања укључује анализу података кроз време. Подаци се дефинишу као хронолошки низ изабраног параметра. У нашем случају је то цена акције. Анализирањем овог низа желимо да уочимо нека правила, периоде, циклусе, трендове у подацима. У нашем случају унапред знање цене акције би помогло у инвестирању пара паметно.

Предвиђање на берзи је деценијама било предмет проучавања и истраживања. То је веома атрактивна тема како инвеститорима тако и истраживачима. Поставља се питање зашто после толико времена нико није успео да реши овај проблем. Одговор је велика комплексност и хаотичност тржишта. Број параметара и извора информација потребних за предвиђање је превелик. Неки верују у хипотезу ефикасног тржишта, по којој се акције увек продају по фер вредности у том тренутку, што би побеђивање тржишта учинило немогућим. Међутим, људи нису рационални, стога многи верују да је могуће предвидети будуће цене акција иако се проблем показао изузетно тешким. Постоји много радова из различитих области који се суочавају са овим изазовом.

Постојеће методе за предвиђање цене деоница се могу поделити на: фундаменталну анализу, техничку анализу и анализу хронолошког низа података. Фундаментална анализа је врста инвеститорске анализе где се будућа вредност деонице процењује помоћу продаја, зараде, профита и других економских фактора. Овај начин се највише користи за дугорочно предвиђање. Техничка анализа разматра историјску цену акција. Пример техничке анализе је померајући просек (moving average). Она се користи за веома кракторочно предвиђање. Последња, анализом хронолошког низа цена деоница, за разлику од техничке анализе, не посматрају се само подаци него и њихове промене кроз време и дају им се тежине (нису све промене исто битне). Дели се на линеарне и нелинеарне алгоритме. Ми ћемо се фокусирати на нелинеарне, прецизније дубоке неуронске мреже. Оне се могу посматрати као контејнери за неку нелинеарну функцију.

Постоји велики број архитектура неуронских мрежа у зависности од примене. Рекуренте неуронске мреже су се показале најмоћније у процесирању секвенцијалних података. Такве мреже поседују кракторајно "памћење". Баш због те способности се овакав модел најчешће покаже боље од стандардних приступа машинског учења, када је у питању овакав тип проблема.

2 Хипотеза

Користићемо LSTM (long short-term memory). Ова архитектура је једна од најуспешнијих RNN-ова. LSTM за разлику од класичних рекурентних неуронских мрежа поседује и дугорочно "памћење" које реализује тако што "забравља" меморију коју сматра ирелевантном за предвиђање следећег излаза. Из тог разлога је способнији да учи из дугих низова података. Ако постоје неки дугорочни трендови у хронолошком низу цена акција на берзи, LSTM ће најбоље моћи да их научи и даће најбоље резултате.

Други метод који ћемо користити и поредити са LSTM-ом је CNN (convolutional neural network). Овај метод би могао да покаже чак и боље резултате од LSTM-а, иако LSTM даје најбоље резултате у многим другим анализама хронолошких низова података.

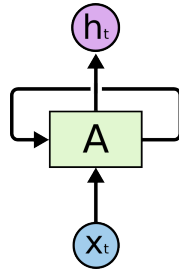
Мотивација иза коришћења ова два модела је да би се видело да ли у променама цена деоница постоје неке дугорочне зависности. Претпоставка је да је берза веома непредвидива, иако има неке трендове и патерне они се мењају током времена и у зависности од компаније и њеног сектора. То би могло да погорша резултате LSTM-а, док CNN користи податке дате у овом тренутку за предвиђање, па би анализа тренутних трендова дала боље резултате.

3 Метод

3.1 RNN

Feedforward неуронске мреже (као нпр. CNN) сликају улаз фиксираних величина у излаз. Веома су моћне и користе се за много различитих проблема. Међутим, доста података није у форми вектора фиксираних величина, већ у форми секвенци. Овај проблем је решен коришћењем рекурентних неуронских мрежа (RNN). RNN је дизајниран тако да може да учи из секвенцијалних података. Људи не почињу да размисљају испочетка сваки пут кад добију нове информације. Користимо се претходним знањима да бисмо разумели наредне информације. За разлику од рекурентних, класичне неуронске мреже то не могу. Помоћу неке врсте "меморије", мрежа се тренира да запамти претходни улаз чиме храни будуће излазе. RNN је природно морао да настане да би могли да се носимо са различитим врстама секвенцијалних података као што је звук, видео, текст.

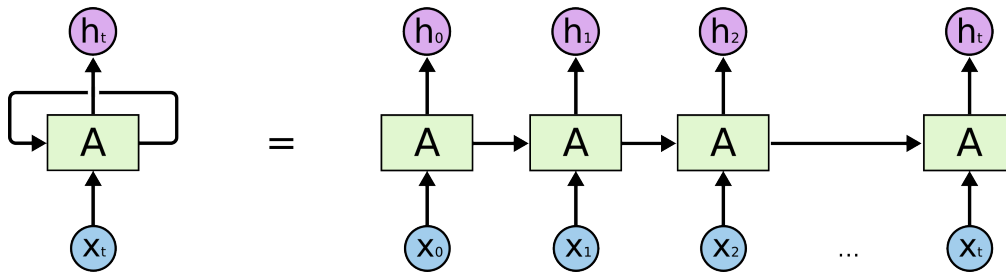
Сада ћемо видети како заправо RNN успева да искористи концепт памћења (веома слично како је заправо реализовано краткотрајно памћење у људима). RNN укључује петље које омогућавају информацијама да опстану.



Слика 1: Рекурента неуронска мрежа приказана са петљом

На слици изнад видимо да се део неуронске мреже означен са A храни са неким улазом x_t и претходним излазом h_{t-1} и даје као излаз h_t . Ова петља даје могућност мрежи да пренесе информацију са једног корака на други.

Ове петље могу изгледати збуњујуће. Међутим, RNN може да се посматра као обична неуронска мрежа, ако се уместо петље мрежа прикаже као мноштво копија једне исте мреже и свакој следећој се прослеђује излаз претходне (одмота).



Слика 2: Одмотана рекурента неуронска мрежа

На слици изнад интуитивно видимо да је рекурента мрежа заправо уско повезана са низовима и секвенцама. Природно је користити такве мреже када се бавимо таквом врстом података.

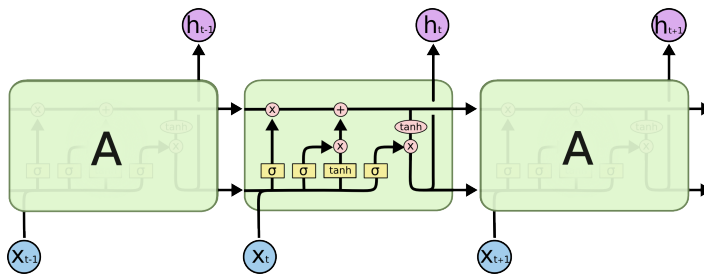
У стандардним рекурентним неуронским мрежама у скривеном стању (на слици означеном са A) се налази само један слој неурона са \tanh активационом функцијом.

У теорији RNN је у могућности да повеже било коју претходну информацију са тренутним проблемом. Нажалост, то није случај и у пракси. Ако је информација превише удаљена почиње да се дешава појава која се назива нестајући градијенти, односно градијенти крену да теже нули. Овај проблем је решен једном специфичном структуром рекурентне неуронске мреже: LSTM.

3.2 LSTM

LSTM (long short term memory) је специјалан тип RNN-а који има могућност да научи и дугорочне зависности. LSTM функционише веома добро на разноврсним проблемима и тренутно је већина водећих модела управо LSTM. Конструисан је да реши проблем "заборављања", штавише дугорочно памћење је дифолтно понашање ове архитектуре. LSTM

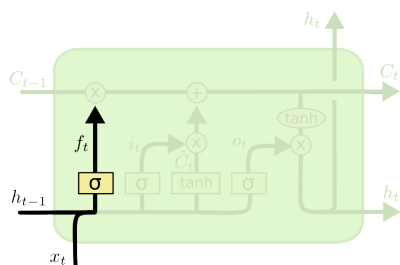
такође има структуру ланца као класичан RNN, али је скривено стање комплексније од једног слоја са \tanh функцијом. У њему се налази четири слоја који интерагују на специјалан начин.



Слика 3: LSTM ланац и скривено стање, означено са A

Главна идеја овог модела је стање ћелије, односно хоризонтална линија која се налази на врху дијаграма. Стање ћелије је преносник информација, који се креће кроз целу структуру са неким малим, линеарним променама (као што су множење и сабирање приказани љубичастим кругом на слици). То омогућава информацијама да се крећу непромењене. LSTM има могућност да додаје или обрише информације са стања ћелије и то контролисано помоћу структура, које се зову капије. Оне су сачињене од слоја неуронске мреже са сигмоид активационом функцијом и операцијом множење. Сигмоид слој враћа бројеве између 0 и 1, што одређује колико ће се неке информације пропустити. LSTM има три овакве капије.

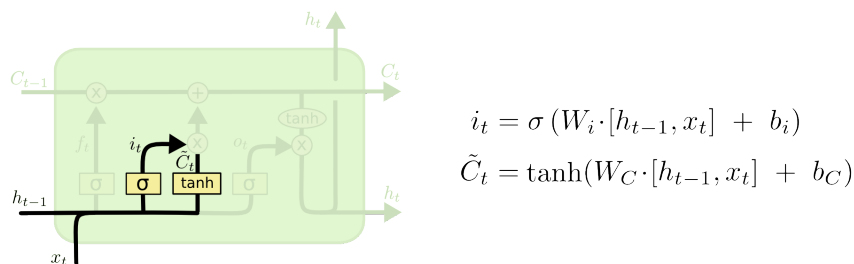
Прва ствар коју LSTM ради је да одлучи шта ће да заборави, тј. које информације ћемо да обришемо из стања ћелије. То се дешава у првом сигмоид слоју. Тај слој се храни са h_{t-1} , претходним излазом и x_t , новим улазом и враћа број између 0 и 1 за сваки број у стању ћелије C_{t-1} . То се множи са стањем ћелије.



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

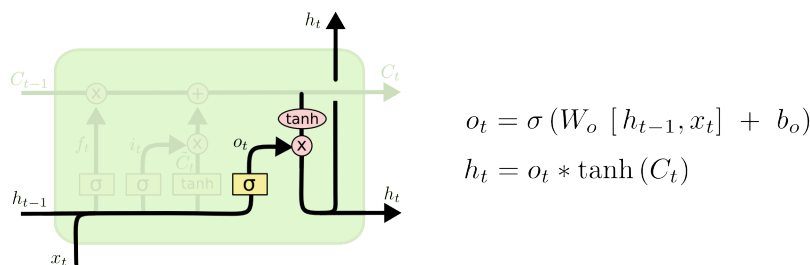
Слика 4: Сигмоид слој задужен за заборављање, f_t излаз, W_f и b_f тежине грана и bias-и

Следећи корак нам је да одлучимо коју нову информацију треба да "запамтимо", тј. да сачувамо у стању ћелије. Прво, сигмоид слој одлучује које вредности ћемо да променимо. После тога \tanh слој конструише вектор кандидата који ће се можда додати у стање ћелије. Следеће множењем та два добијамо вредности које ћемо додати на стање ћелије.



Слика 5: Део архитектуре задужен за додавање нових информација на стање ћелије

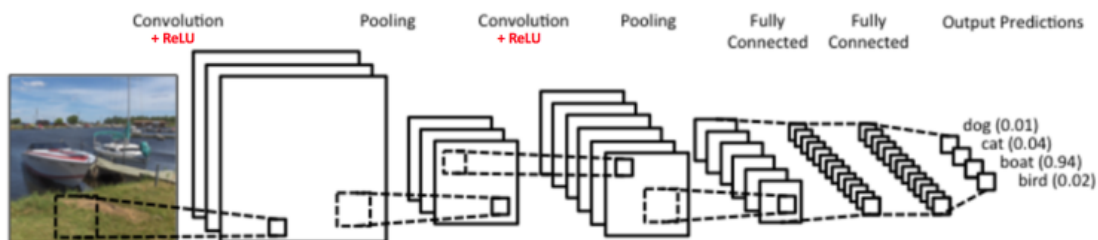
После ова два корака добијамо ново стање ћелије C_t . Сада можемо одлучити и шта ћемо да вратимо. Излаз је базиран на стању ћелије. Улаз и претходни излаз пролазе кроз сигмоид слој, који одлучује које ћемо заправо делове да вратимо. Након тога стање ћелије пролази кроз \tanh слој и множи се са излазом сигмоид слоја да би вратили само жељене делове.



Слика 6: Последњи корак пре враћања излаза

3.3 CNN

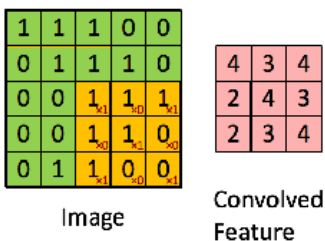
CNN (convolutional neural network) је "feed-forward" неуронска мрежа. Први пут је конструисана 1988 и звала се LeNet. Конволуционе неуронске мреже се најчешће користе за препознавање слика, из разлога што су веома добре у учењу одређивања feature-а тј. карактеристика слике. То би могло да се покаже добро и на проблему предвиђања цене акција. Конволуционе мреже имају четири најбитније операције: конволуција, нелинеарност, pooling, потпуно повезан слој.



Слика 7: Архитектура прве конволуционе неуронске мреже LeNet

3.3.1 Конволуциони слојеви

Име конволуциони слојеви и саме мреже, конволуциона неуронска мрежа, изведено је из имена математичке операције конволуција. Главна примена конволуције у CNN-у је да одреди feature-e. Конволуција се извршава тако што кернел (матрица чије димензије ми одредимо) прође кроз другу матрицу (већих димензија) и свако поље кернела се множи са одговарајућим пољем матрице.



Слика 8: Конволуција, лево је матрица над којом се ради конволуција, десно је новонастала матрица, жутим је приказан кернел

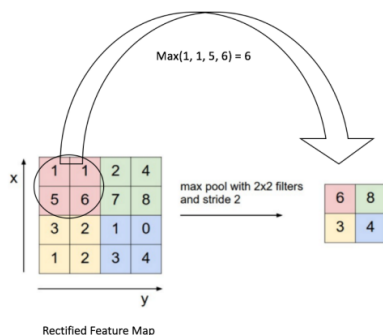
У примеру на слици померамо кернел за једно поље у матрици, али постоји могућност да се помера и за више. Тај померај се зове корак (stride).

Користећи терминологију за неуронске мреже кернел се назива филтер и дефинишу се његове димензије. Током тренирања вредности у филтерима се уче (оптимизују) како би што боље одређивали feature-e, али се величина и број филтера предефинише. Величина следећег слоја зависи од броја филтера, тиме се одређује дубина (број канала). Даље корак (stride) одређује величину сваке матрице. Такође постоји могућност попуњавања, додавања поља као оквир (и то може да утиче на величину следећег слоја). То се ради јер иначе не може кернел да дође до ћошкова или да бисмо повећали матрицу која настаје.

3.3.2 Pooling слој

Pooling операција се користи да би се смањиле димензије сваког канала који се добије после конволуционог слоја, али да се уједно задрже најбитније информације. Постоји различити типови pooling-a: max, просек, сума.

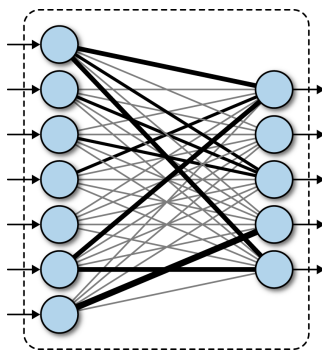
У случају max pooling-a дефинишемо величину околине (на пример прозор 2x2) и узимамо највећи елемент из матрице који се налази у том прозору.



Слика 9: Max pooling операција

3.3.3 Потпуно повезан слој

Последњи корак у конволуционим неуронским мрежама је потпуно повезан слој. Потпуно повезан означава да је сваки неурон у претходном слоју повезан са сваким неуроном у следећем слоју. Излаз конволуционих и pooling слојева ће бити представљен са великим бројем канала. Из тог разлога се прво све информације смештају у једнодимензионални вектор, после чега се потпуно повезује са наредним слојем. Потпуно повезан слој нам враћа жељене информације из одређених feature-a почетног улаза.



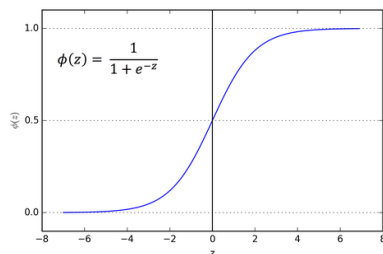
Слика 10: Потпуно повезан слој

3.4 Активационе функције

Активационе функције су веома битне у неуронским мрежама да не би дошло до "експлодирања" вредности неурона. Како су неуронске мреже веома комплексне и много бројева се сабира и множи, потребно је да компресујемо бројеве на неки мањи интервал. Активационе функције које користимо су сигмоид, tanh и ReLu.

3.4.1 Сигмоид

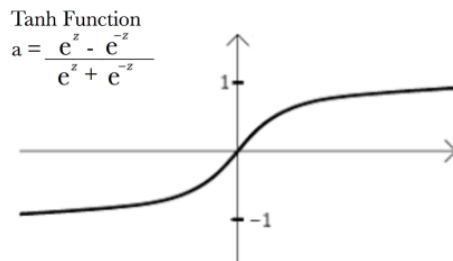
Сигмоид функција је веома погодна јер пресликава бројеве у интервал од 0 до 1. Најчешће се користи за моделе где треба да се предвиди нека вероватноћа (како је вероватноћа увек између 0 и 1). У LSTM-у нам је потребна за капије, јер одређујемо колико нешто треба да се заборави или дода. Нула значи потпуно заборави или немој уопште да додајеш, док један означава да се сачува нека информација или потпуно дода.



Слика 11: Сигмоид активациона функција

3.4.2 Tanh

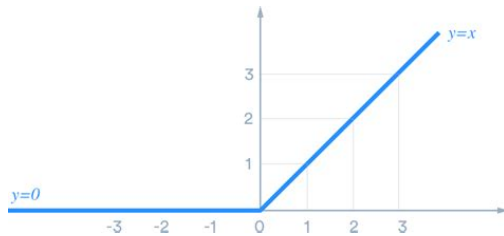
Tanh (хиперболички тангенс) је веома слична функција као и сигмоид, само што пресликава у интервал од -1 до 1. Такође је у S облику као сигмоид.



Слика 12: Tanh активациона функција

3.4.3 ReLU

Активациона функција која се користи за конволуционе слојеве је ReLU. ReLU је скраћено за Rectified Linear Unit. Она је веома битна у CNN-у јер уводи нелинеарност коју смо спомињали. Она је дефинисана као $f(x) = \max(0, x)$. За конволуционе слојеве се најбоље показала ReLU активациона функција.



Слика 13: ReLU активациона функција

3.5 База података

Историјски подаци за деонице се могу преузети на сајту Yahoo Finance (<https://finance.yahoo.com>) у CSV (Comma Separated Values) формату. То је текст датотека која садржи листу података раздвојених зарезом. За рад са подацима користимо python библиотеку pandas. Користе се историјски подаци компаније Apple Inc. (AAPL) од 1.1.2000. па до 20.5.2021. (5380 одабирака) За сваки датум имамо 6 вредности: open, close, high, low, adj. close, volume. Радимо са close вредности деонице, тј. цена деонице на затварању трансакција тог дана.

База се препроцесира. Ради се нормализација на подацима и подаци се трансформишу у интервал од 0 до 1.

Након тога база је подељена на тренинг (4304 одабирака, 80% базе) и тест (1076, 20% базе) скупове, да бисмо могли да евалуирамо моделе.

3.6 Модели

Као што је објашњено CNN се састоји од два битна слоја, конволуциони и pooling. Конволуциони слојеви могу да имају улаз у форми дводимензионалне слике или као низ једнодимензионалних података. У нашем случају је друго наведено. Резултати који се добију после конволуционих слојева се пројектују у feature mapу. После тога над тиме раде pooling слојеви, да би се одредили најбитнији feature-и.

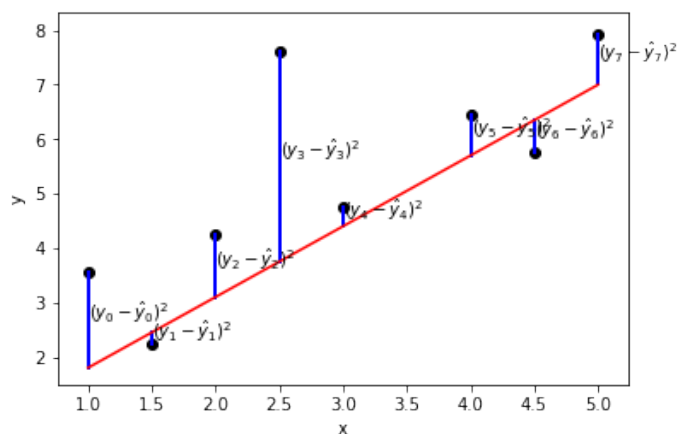
Први модел је CNN који као улаз користи претходних 50 дана података. Тренирамо модел користећи податке из скупа за тренирање. Предвиђамо цену при затварању трансакција наредног дана. Дакле улаз је димензија (50, 1). Модел је сачињен од само једног конволуционог слоја, који одређује 16 feature мапа, односно има 16 филтера величине 3. После pooling слој изводи max pooling величине 2, смањујући димензије feature мапа. На крају се то изравна у једнодимензионални вектор што је улаз за следећи потпуно повезани слој са 25 неурона, који је пре излазног слоја који предвиђа наредан дан, тј. има 1 неурон. ReLU се користи у конволуционим и потпуно повезаним слојевима. Пеформансе модела се оптимизују коришћењем ADAM верзије стохастичког gradient descent-а. Модел се тренира са 20 епоха са batch-евима величине 50. За функцију губитка узет је MSE (mean square error) који представља аритметичку средину квадрата разлика предвиђене и реалне вредности.

Други модел је LSTM који се састоји из енкодер и декодер подмодела. Користи претходних 50 дана као улаз. Енкодер се користи да прочита и енкодује улазни низ док декодер треба да интерпретира енкодирану секвенцу. Првом LSTM слоју се предаје улаз димензија (50, 1), односно претходних 10 недеља података. Он је сачињен од 50 чворова. Излаз овог LSTM слоја је вектор са 50 елемената (један по чвору) који одређује feature-е од улазних 50 вредности. Сада се то декодира са још једним LSTM слојем са такође 50 чворова. Даље се то шаље потпуно повезаном слоју са 25 неурона пре него што дође до излазног слоја са 1 неуроном. MSE је функција губитка. ADAM је оптимизатор. Тренира се на 20 епоха са величином batch-а 50.

4 Метрика

За анализирање ефикасности модела користимо Root mean squared error (RMSE). RMSE је корен просечне вредности квадрата свих грешака, односно разлика стварне и предвиђене вредности. Коришћење RMSE је веома често и представља одличну генералну метрику за нумеричка предвиђања. У поређењу са средњом апсолутном грешком, RMSE појачава и строго кажњава велике грешке.

$$RMSE = \sqrt{\frac{1}{N} \sum (\hat{Y}_i - Y_i)^2} \quad (1)$$

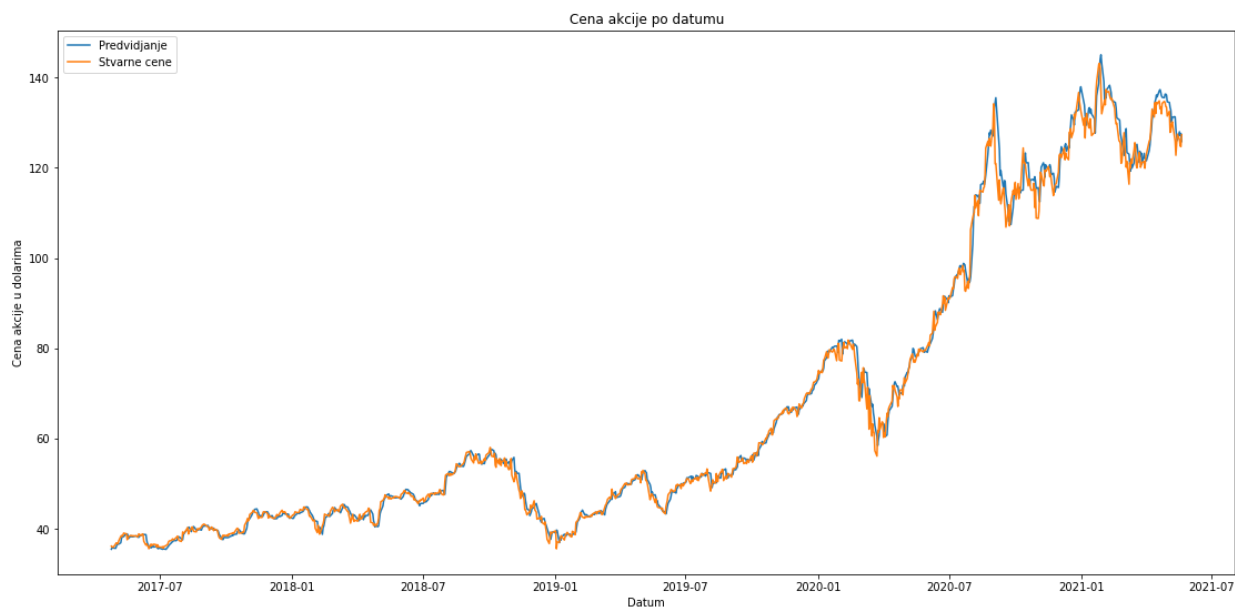


Слика 14: RMSE

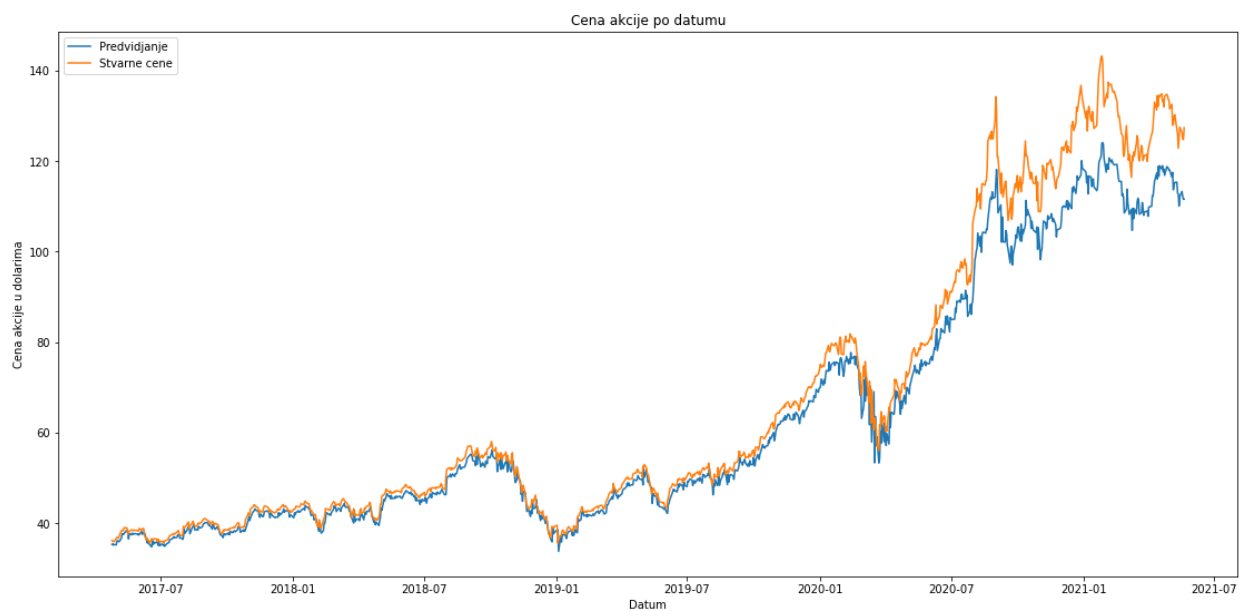
5 Резултати

Модели су тестирани на тест скупу података, тј. на подацима од 2017. па до 2021. Најбоље се показао CNN модел са 200 епоха, чији RMSE је био 2.07653.

Модел	Број епоха	RMSE
CNN	20	4.20347
	50	2.96310
	100	2.28429
	200	2.07653
	300	2.11713
LSTM	20	13.85259
	50	8.01862
	100	6.45549
	200	10.59045
	300	19.78913



Слика 15: График предвиђених и стварних вредности деоница компаније Apple Inc. CNN модела тренираног са 200 епоха



Слика 16: График предвиђених и стварних вредности деоница компаније Apple Inc. LSTM модела тренираног са 100 епоха

6 Закључак

Популарност инвестирања на берзи константно расте, што подстиче истраживаче да проналазе нове методе за предвиђање цена акција коришћењем нових техника. Да би се помогло инвеститорима потребан је прецизан модел. У овом раду смо предложили два модела у ову сврху: LSTM и CNN. Хипотеза је оправдана, CNN се показао боље. Са претходна два графика јасно се види да су предвиђања модела почела знатно да одступају од реалних вредности од када је почела пандемија COVID-19 вируса, док су до тада веома слично предвиђали. Тада су почеле веће флукуације на које LSTM није могао довољно брзо да се навикне, док је насупрот њему CNN радио подједнако добро и у том периоду. Овај резултат лено илуструје могућност CNN-а да препознаје тренутне трендове.

Литература

- [1] JSidra Mehtab, Jaydip Sen, *Stock Price Prediction Using CNN and LSTM Based Deep Learning Models*, (2010).
- [2] Sidra Mehtab, Jaydip Sen, *Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Time Series*, (2020).
- [3] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, *Predicting Stock Prices Using LSTM*.
- [4] David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira, *Stock Market's Price Movement Prediction With LSTM Neural Networks*.
- [5] Jreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P, *STOCK PRICE PREDICTION USING LSTM,RNN AND CNN-SLIDING WINDOW MODEL*, (2018).
- [6] Yuqing Dai, Yuning Zhang, *Machine Learning in Stock Price Trend Forecasting*.
- [7] J. Sen and T. Datta Chaudhuri, *A robust predictive model for stock price forecasting*, (2017).
- [8] Murtaza Roondiwala¹, Harshal Patel, Shraddha Varma *Predicting Stock Prices Using LSTM*, (1997).
- [9] Sepp Hochreiter, Jürgen Schmidhuber *Long Short-term Memory*, (1997).