

Предикција цена лаптопова на основу карактеристика

Anonymised Author

Софтверско Инжењерство и Информационе Технологије, Универзитет у Новом Саду, Факултет техничких наука
surname.sv.xx.yyyv@uns.ac.rs

I. ПРОБЛЕМ

Иако су од самог почетка свог постојања рачунари представљали важан алат у људском животу, у данашњем све напреднијем технолошком окружењу постали су незаменљиви уређаји како за појединце, тако и за предузећа. Преносивост и свестрана примена лаптопова довели су до значајног раста базе корисника, а тржиште лаптопова почело је да нуди широк спектар производа са различитим спецификацијама и ценама, што потрошачима отежава информисање и доношење разумних одлука приликом куповине [1]. Глобално тржиште лаптопова достигло је вредност од приближно 259,11 милијарди америчких долара у 2024. години и очекује се да ће наставити да расте по просечној годишњој стопи раста од око 8,8% у периоду од 2025. до 2028. године [2]. На ову статистику и пораст потражње за лаптоповима додатно је утицала и пандемија Ковида-19 услед преласка на рад од куће [3].

Циљ овог пројекта је да развије модел машинског учења који предвиђа цене лаптопова на основу њихових карактеристика, као што су марка, тип процесора, димензије екрана, резолуција, РАМ меморија, капацитет складишта, графичка картица (ГПУ), оперативни систем и тежина и да анализом скупа података који садржи поменуте спецификације и цене идентификује кључне факторе који утичу на формирање цене. Како би се одредила разумна цена лаптопа, у литератури су предложени различити модели предвиђања, као што су методе класификације засноване на алгоритмима стабала одлучивања, као и различити регресиони модели, уз истовремену примену регресионих метрика за процену тачности предикција и идентификацију најпрецизнијег модела [4].

Резултати овог пројекта пружиће вредне увиде како потрошачима, тако и трговцима на тржишту лаптопова. Модел ће потрошачима омогућити процену фер цене лаптопа према изабраним карактеристикама, а трговцима ће помоћи у оптимизацији ценовних стратегија и задржавању конкурентске позиције.

На крају, овај пројекат унапређује транспарентност и оптималност тржишта лаптопова јер потрошачима омогућава да лакше процене вредност производа и донесу промишљене одлуке, док продавцима пружа основу за конкурентније и праведније формирање цена.

II. ТЕОРИЈСКЕ ОСНОВЕ

У овом поглављу представљају се теоријске основе потребне за разумевање проблема предикције цена лаптопова и одабраног решења. Организација поглавља подељена је у два потпоглавља. У потпоглављу II.A појашњава се детаљније сам домен проблема и идентификовани захтеви које решење треба да испуни, док се у II.B разлажу основни концепти решења и примењене методе машинског учења.

A. Домен и захтеви система

Предикција цена лаптопова се заснива на регресионој анализи, где је циљ да се на основу техничких карактеристика уређаја као што су процесор, РАМ меморија, капацитет складишта и бренд предвиди његова тржишна цена. Ово решење би имало широку примену у доношењу одлука о куповини, продаји и креирању стратегија цена на тржишту.

Постојећа решења се често ослањају на ручне анализе или основне статистичке методе које не могу у потпуности да обухвате сложене релације између више фактора, што доводи до нетачних предикција и ограничене поузданости [5]. На основу циљева интересних група, у Табели 1 издвојени су захтеви које систем треба да испуни.

B. Имплементација решења

Претходно припремљени подаци су обрађени и спремни за коришћење у моделима. Укратко, извршено је чишћење података, кодирање категоријалних променљивих и нормализација нумеричких вредности, како би се обезбедила доследност и квалитет података.

Након припреме скупа података, приступило се имплементацији алгоритама машинског учења. Изабрани су алгоритми који се у пракси често показују ефикасним у регресионим задацима: *Ridge*, *Decision Tree*, *Random Forest* и *Extra Trees*. Сваки од ових модела пружа различите предности у зависности од природе података и комплексности задатка.

За сваки модел спроведена је оптимизација хиперпараметара методом грид претраге, уз примену *5-fold* крос валидације ради поузданије процене перформанси и смањења ризика од *overfitting*-а модела, а са циљем добијања што бољих резултата предвиђања. Као метрика перформанси у фази евалуације коришћена је средња апсолутна грешка (MAE), јер она јасно показује просечно одступање предвиђене од стварне цене као и корен средње квадратне грешке (RMSE).

Поређењем добијених резултата, идентификован је модел који има најбољи однос између тачности и сложености. Овим приступом омогућено је да систем не само предвиђа цене што прецизније, већ и да задржи стабилност и поузданост у различитим условима. оквир који омогућава разумевање и поређење различитих приступа регресији.

Табела 1 - Апликативна решења

Апликативни о решења	Интересна група	Циљ крајњег корисника	Захтев система
Аутоматска процена вредности лаптопа	Купци, продавци, сервисни центри	Брза и поуздана процена реалне тржишне цене лаптопа	Модел мора да обрађује улазне податке и да генерише тачну цену у реалном времену
Поређење лаптопова према вредности	Купци, студенти, ИТ ентузијаста	Избор најбоље понуде на тржишту у складу са буџетом и карактеристикама	Систем мора подржавати упоређивање више уноса истовремено
Подршка одлукама приликом куповине или продаје	Купци и продавци	Добра информисаност пре доношења одлуке о куповини или препродаји	Интерфејс мора бити интуитиван и резултати разумљиво приказани
Анализа тржишних трендова	Продавци, маркетинг тимови, аналитичари	Добијање увида у кретање цена током времена	Систем мора чувати и анализирати историјске податке, као и пружати визуелне и графичке приказе резултата
Оптимизација стратегија цена за продавце	Продавци, онлајн продавнице	Ефикасно одређивање цена које максимизују профит	Модел мора подржавати различите сценарије и прогнозе

III. РЕШЕЊЕ

Циљ решења је развој модела машинског учења који предвиђа цену лаптопа на основу његових спецификација, као што су марка, тип процесора, величина и резолуција екрана, РАМ меморија, капацитет складишта, графичка картица, оперативни систем и тежина.

Улаз у систем представља скуп података са спецификацијама и постојећим ценама лаптопова, док је излаз процењена цена новог уређаја. На [Слици 1](#) приказан је дијаграм процесног тока решења које је организовано у неколико фаза (А-Е):

А. Прикупљање и припрема података

За израду решења коришћен је скуп података који садржи спецификације и цене лаптопова, учитан помоћу библиотеке *pandas* са платформе *Kaggle* [6]. Пре почетка анализе било је неопходно обезбедити јединствен формат података. Анализом података уочено је да постоје недостајуће вредности у колонама као што су РАМ и складиште. Недостајући подаци за РАМ попуњени су просечном вредношћу у оквиру типа лаптопа, док је за тип складишта коришћена најчешћа вредност. Колоне које имају велики број недостајућих вредности и малу важност за анализу (нпр. *ID* модела) уклоњене су из скупа података.

Категоријске колоне (марка, процесор, оперативни систем) трансформисане су коришћењем *one-hot encoding* и *label encoding* техника. Нумеричке колоне су стандардизоване коришћењем *StandardScaler*-а како би се омогућила упоредивост вредности и боља конвергенција модела.

Б. Екстракција обележја

Дефинисани су кључни параметри који највише утичу на цену лаптопа: број језгара процесора, капацитет РАМ-а, тип складишта, марка и оперативни систем. Генерисане су и нове особине по потреби, на пример однос цене по гигабајту меморије и укупан капацитет меморије, како би модел имао додатне информације за предикцију.

Излаз овог корака је вектор обележја за сваки лаптоп, спреман за тренирање регресионних модела.

Ц. Подела скупа и тренирање модела

Подаци су подељени на тренирани и тест скуп у односу 80:20. За анализу су примењени различити регресионни модели: *Ridge*, *Decision Tree*, *Random Forest* и *Extra Trees*.

Такође је коришћен *GridSearchCV* са *5-fold cross-validation* за оптимизацију хиперпараметара сваког модела, што омогућава проналажење најбоље комбинације параметара и смањење ризика од преоптерећења (*overfitting*) [7].

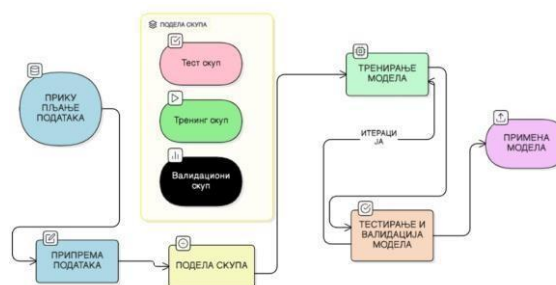
Д. Евалуација модела

Циљ евалуације је утврдити који модел даје најтачније предикције цена лаптопова на непознатим подацима. Пожељни исходи су минималне вредности метрика средње апсолутне грешке (*MAE*) и квадратног корена средње квадратне грешке (*RMSE*), као и што већа вредност коефицијента детерминације (*R²-score*) [8].

Анализа важности обележја омогућава увид у то који параметри највише утичу на цену, што може бити корисно за даљу оптимизацију и пословне одлуке.

Е. Примена модела

Крајњи корисник може унети карактеристике лаптопа, а систем враћа процењену цену која представља оптималну вредност на основу тржишних података и модела.



Слика 1 - Дијаграм процесног тока решења

IV. РЕЗУЛТАТИ И ДИСКУСИЈА

У овом поглављу приказани су и објашњени резултати тестирања и евалуације у контексту предикције цена. Циљ је био да се испита тачност и поузданост различитих регресионних модела и идентификује онај који најбоље предвиђа цене.

За анализу су примењени следећи модели: *Ridge* регресија, која уводи регуларизацију и смањује *overfitting*; *Decision Tree*, који користи структуру стабла за доношење одлука; *Random Forest*, ансамбл метода која комбинује више стабала и повећава тачност; и *Extra Trees*, сличан *Random Forest*-у али са додатном насумичношћу при поделама. Сви модели су тренирани на 80% скупа података и тестирани на 20%, уз мерење перформанси помоћу *MAE* (средња апсолутна грешка), *RMSE* (средња квадратна грешка) и *R²-score*.

Да би се побољшале перформансе модела и смањила грешка предикције, за све алгоритме је примењена техника *GridSearch*-а у комбинацији са *k-fold* крос-валидацијом (*k*=5). Оптимизовани су следећи хиперпараметри:

- *Ridge* регресија: *alpha* (интензитет регуларизације - веће вредности значе јачу контролу сложености модела)
- *Decision Tree*: *max_depth* (највећа дозвољена дубина стабла - ограничава прекомерно учење), *min_samples_split* (минималан број узорака потребан да би се чвор поделио), *min_samples_leaf* (минималан број узорака који морају остати у листу стабла)
- *Random Forest*: *n_estimators* (број стабала у шуму - више стабала обично повећава стабилност предвиђања), *max_depth*, *min_samples_split*, *min_samples_leaf*.
- *Extra Trees*: *n_estimators* (број стабала у ансамблу), *max_depth*, *min_samples_split*, *min_samples_leaf*.

Табела 2 - Оптимизовани хиперпараметри

Модел	Најбољи хиперпараметри
Ridge регресија	<i>alpha</i> = 10
Decision Tree	<i>max_depth</i> = 15, <i>min_samples_split</i> = 5, <i>min_samples_leaf</i> = 2
Random Forest	<i>n_estimators</i> = 200, <i>max_depth</i> = 20, <i>min_samples_split</i> = 5, <i>min_samples_leaf</i> = 2
Extra Trees	<i>n_estimators</i> = 300, <i>max_depth</i> = 25, <i>min_samples_split</i> = 4, <i>min_samples_leaf</i> = 1

Вредности параметара приказаних у Табели 2 дале су најбоље резултате након вишеструких тестирања и оптимизације, при чему је циљ био минимизовање *MAE* и *RMSE* уз максималан *R²-score*.

Табела 3 - Евалуација решења

Модел	MAE (евра)	RMSE (евра)	R ² -score
Ridge регресија	169.07	220.45	0.87
Decision Tree	185.32	245.67	0.81
Random Forest	171.88	218.12	0.86
Extra Trees	174.56	221.09	0.85

Из Табеле 3 је видљиво да *Ridge* регресија постиже најбоље резултате са најнижом средњом апсолутном грешком (*MAE* = 169.07 €) и највишим *R²-score* (0.87), што указује на високу тачност предвиђања. *Random Forest* има врло сличне резултате, што потврђује стабилност ансамбл метода. *Extra Trees* даје благо слабије резултате због веће насумичности у подели података, док је *Decision Tree* најмање прецизан, што указује на тенденцију *overfitting*-а када се користи самостално.

Анализа важности карактеристика показала је да највећи утицај на цену имају: марка лаптопа, тип процесора, РАМ меморија и капацитет складишта. Ови фактори највише објашњавају варијацију у ценама, што је у складу са очекивањима тржишта хардвера.

Резултати су стабилни, али зависе од квалитета скупа података. Недостајуће вредности су захтевале импутацију, што може унети шум. Такође, модели су тренирани на одређеном тржишном сегменту, па је њихова генерализација на друге скупове ограничена. У будућем раду могу се укључити додатне карактеристике као што су трајање батерије, тежина или резолуција екрана, као и напреднији ансамбл модели или неуронске мреже.

V. ЛИТЕРАТУРА

- [1] Christy Andrews J, Kavinesh V S, LAPTOP PRICE PREDICTION, International Research Journal of Education and Technology, ISSN 2581-7795
- [2] The Business Research Company, Laptops Global Market Report 2024.
- [3] Historic laptop demand leads to shortages ahead of remote school, Axios, August 15, 2020.
<https://www.axios.com/2020/08/15/coronavirus-education-laptop-shortage> [13. јул 2025]
- [4] Gulzat Turken, Lyazat Naizabayeva, Siládi Vladimir, et al. RESEARCH ON PREDICTIVE MODEL BASED ON CLASSIFICATION WITH PARAMETERS OF OPTIMIZATION. Neural Network World, 2020 (5): pp. 295-308
- [5] Prof. B. Prajna, A. Lekha Sri, A. Sahiti, A. Bhagyasri Teja, A. Arzoo LAPTOP PRICE PREDICTION USING MACHINE LEARNING, Department of CSE, AUCEW, Visakhapatnam
- [6] Скуп података: <https://www.kaggle.com/datasets/jacksondivakarr/laptop-price-prediction-dataset>
- [7] *GridSearchCV* - *scikit-learn* документација https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
- [8] *Scikit-learn* - *Regression metrics* https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics