



**Departamenti Shkenca Kompjuterike**

# **Ndikimi i Machine Learning në parashikimin e shitjeve të produkteve**

Niveli Bachelor

Albion Burniku

Qershor/2023

Ferizaj



**Departamenti Shkenca Kompjuterike**

Punim Diplome

2020/2021

Albion Burniku

# **Ndikimi i Machine Learning në parashikimin e shitjeve të produkteve**

Mentor: Prof.Dr. Hyrmet Mydyti

Qershor/2023



## **Abstrakti**

Objektivi ynë në këtë studim ishte të identifikonim teknikat më efektive që ndikuan në shitjet e produkteve përmes algoritmeve të mësimi të makinerive. Duke pasur parasysh faktorë të ndryshëm si çmimi i produktit, ofertat promovionale dhe vlerësimet e klientëve, u përcaktua se si këto variabla ndikuan në performancën e tregut.

Vlerësimet e klientëve u gjetën të kenë efektin më të madh në shifrat e shitjeve me çmimet dhe ofertat promovionale që kanë më pak ndikim në krahasim. Për të parashikuar me saktësi performancën e produktit në të ardhmen duke përdorur grupet aktuale të të dhënave, ne përdorëm gjuhën e programimit Python në lidhje me algoritmet e machine learning si regresioni linear, regresioni i kreshtës, regresioni i decision tree dhe regresioni i rëndom forest.

Gjetjet tona treguan se algoritmi i decision tree regressor kishte shkallën më të lartë të saktësisë kur parashikonte gjenerimin e të ardhurave të ardhshme nga produktet. Pasojat e këtij studimi sugjerojnë se bizneset mund të përfitojnë nga përmirësimi i vlerësimeve të përgjithshme të klientëve, ndërkohë që ofrojnë njëkohësisht struktura çmimesh konkurruese për të arritur rezultate më të mira të shitjeve në të ardhmen.

**Fjalët Kyçe:** Machine Learning, Parashikim, Python, Algoritme, Shitje

## **Abstract**

Our objective in this study was to identify the most effective techniques that impacted product sales through machine learning algorithms. Given various factors like product price, promotional offerings, and customer ratings, it was determined how these variables affected market performance. Customer reviews were found to have the biggest effect on sales figures with prices and promotional offers having the least impact on comparison. To accurately predict product performance in the future using current data sets, we used The Python programming language in convening with machine learning algorithms like linear regression, crest regression, decision tree regression, and random forest regression. Our findings showed that the decision tree regressor algorithm had the highest degree of accuracy when predicting future revenue generation from products. The implications of this study suggest that businesses may benefit from improving overall customer ratings while simultaneously providing competitive pricing structures to achieve better future sales results.

**Keywords:** Machine Learning, Prediction, Python, Algorithms, Sales.

## **Deklarata e origjinalitetit**

Deklaroj me përgjegjësi të plotë personale unë Albion Burrniku, që punimi i paraqitur, në dijeninë time nuk përfshin material apo të dhëna, të prezantuara më parë në një punim diplome në ndonjë universitet dhe se në dijeninë time nuk përmban ndonjë material të publikuar më parë me shkrim nga ndonjë person tjetër. Çdo material i nevojshëm i përfshirë në tekstin e këtij punimi është referuar sipas rregullave.

---

## **Falënderimet**

Ndihem i obliguar të falënderoj për kontributin e dhënë, familjen time gjatë këtyre tri viteve studime duke mos lënë anash as kolegët dhe shoqërinë time me të cilët kam punuar për vite me radhë.

Një falënderim i veçantë shkon për mentorin tim, Prof.Dr. Hyrmet Mydyti, për ndihmën dhe mbështetjen e çmuar që më ka ofruar përgjatë gjithë punës sime dhe për kontributin e saj në finalizimin e punimit tim të diplomës, produkt i orëve të konsultimit, këshillimit dhe mbështetjen nga ana e saj.

Po ashtu, dua të falënderojë stafin administrativ dhe akademik të Kolegjit Universum për ndihmën e ofruar gjatë secilit vit duke mos hezituar në asnjë rast për të plotësuar kërkesat tona dhe duke na qëndruar afër për çdo vështirësi dhe në çdo moment, gjithashtu edhe për mundësinë e ofruar për të perfunduar eksperiencën në programin Erasmus+.

## Përmbajtja

Abstrakti.....	i
Abstract.....	ii
Deklarata e origjinalitetit .....	iii
Falënderimet .....	iv
Përmbajtja .....	v
Lista e figurave .....	vi
Lista e tabelave .....	vii
Fjalori i termave.....	viii
1. Hyrje .....	1
2. Shqyrtimi i literaturës .....	4
2.1. Teknika Machine Learning .....	4
2.2. Aplikimi i teknologjisë Python në Machine Learning .....	6
2.3. Krahasimi i teknikave të Machine Learning .....	7
3. Metodologjia.....	10
4. Zhvillimi dhe Dizajnimi i Aplikacionit .....	12
4.1. Krijimi i hapësirës punuese .....	12
4.2. Zhvillimi i Parashikimit .....	14
4.3. Zhvillimi i Aplikacionit.....	22
5. Konkluzioni .....	27
6. Rekomandime.....	28
Referencat .....	29
Shtojca A.....	31
Shtojca B.....	33



## Lista e figurave

Figura 1 Machine Learning.....	5
Figura 2 10 Libraritë më të përdorura për Machine Learning .....	6
Figura 3 Pamja e pare e Anaconda .....	12
Figura 4 JupyterLab pamja .....	13
Figura 5 Shitjet totale të produkteve.....	15
Figura 6 Shitjet e produkteve muajin e kaluar .....	16
Figura 7 Shitjet e produkteve javën e kaluar .....	17
Figura 8 Shitjet e produktit gjatë kohës se ditës .....	18
Figura 9 Shitjet e B52 gjatë ditës.....	18
Figura 10 Shitjet VIPA gjatë ditës.....	19
Figura 11 Produktet më të shitura në vitin 2021, 2022 dhe 2023 .....	20
Figura 12 Sidebar për aplikacion .....	23
Figura 13 Zgjedhja e produktit .....	23
Figura 14 Zgjidhni vitin.....	24
Figura 15 Shitjet gjatë orëve të ditës për produktin.....	25
Figura 16 Shitjet e produktit gjatë muajve për vitin e caktuar.....	25
Figura 17 Muaji për më së shumti shitje për produktin.....	26
Figura 18 Shitjet Birra Peja gjatë ditës .....	33
Figura 19 Shitjet Lasko gjatë ditës .....	33
Figura 20 Shitjet DR GERARD gjatë ditës .....	34
Figura 21 Shitjet BRAVO gjatë ditës .....	34

## **Lista e tabelave**

Table 1 Performance of Predictive Models Using 10-Fold Cross-Validation on the Balanced Data Set.....	8
Table 2 Krahاسimi i algoritmeve të ndryshme.....	9
Table 3 Dataseti i perdorur në parashikim.....	11
Table 4 Besueshmëria e algoritmeve .....	22

## **Fjalori i termave**

RF - Random Forest

ML - Machine Learning

IA - Inteligjenca Artificiale

DTR - Decision Tree Regressor

RR - Ridge Regression

LR - Linear Regression

AI - Artificial Intelligence

PD - Pandas

NP - NumPy

PLT - Matplotlib.Pyplot

CMD - Command Prompt

# 1. Hyrje

Parashikimi i shitjeve është një aspekt thelbësor i operacioneve të biznesit pasi ndihmon kompanitë të vlerësojnë shitjet e ardhshme dhe të planifikojnë në përputhje me rrethanat strategjitet e tyre të prodhimit dhe marketingut. Mësimi i makinerisë (ML) është shfaqur si një qasje premtuese për parashikimin e shitjeve pasi mund të shfrytëzojë sasi të mëdha të dhënash dhe të identifikojë modele që mund të mos jenë të dukshme për analistët njerëzorë. Në këtë studim, ne eksplorojmë përdorimin e algoritmeve ML për të parashikuar shitjet e produkteve.

Për t'u dhënë përgjigje pyetjeve tona, ne aplikuam algoritme të ndryshme ML të cilat përfshinin regresionin linear, regresionin e kreshtës, regresionin e pemës së vendimit dhe regresionin e rastësishëm të pyjeve. Arsyeja pas zgjedhjeve tona u informua nga performanca e tyre e kaluar në studime, si dhe nga kapaciteti i tyre për të menaxhuar grupe të gjera të dhënash dhe për të dalluar lidhjet jolineare midis variablave të ndryshëm.

Në dritën e përpjekjeve tona të analizës, u dallua se metoda më e mirë për parashikimin e shitjeve qëndron në regresionin e pemës së vendimit për shkak të shkallës së lartë të saktësisë. Gjetjet tona pasqyrojnë ato të studimeve të kaluara - mes tyre Fernández Delgado et al.s (2021) - të cilat njohën gjithashtu epërsinë e këtij algoritmi mbi të tjerët në këtë fushë.

Për të përcaktuar se cilat strategji kanë ndikimin më të madh në statistikat e shitjeve të produkteve, ne kryem një rishikim të gjerë të faktorëve përkatës. Këto përfshinin strategjitë e çmimeve, marrëveshjet promovues nëpër kanale të ndryshme të shpërndarjes, si dhe reagimet e konsumatorëve si rishikimet në internet ose forma të tjera të mekanizmave të vlerësimit.

Ajo që doli nga ekzaminimi ynë ishte një kuptim i qartë se kur bëhet fjalë për nxitjen e blerjeve përpara në kontekstin e sotëm të tregut - asgjë nuk është më e fuqishme se reagimet pozitive të klientëve për sa i përket fuqisë së ndikimit në rezultate si gjenerimi i të ardhurave ose nivelet e përgjithshme të angazhimit të konsumatorëve.

Në përputhje me studimet e mëparshme (Mohan & Kumar; 2021), analiza jonë konfirmon këtë gjetje duke theksuar gjithashtu se sa e rëndësishme është që markat të vënë theksin më të madh në monitorimin e reputacionit të tyre në internet midis klientëve.

Është thelbësore të përdoret algoritmi i duhur i mësimi të makinerive për parashikim, përveç kërkimit të elementeve që ndikojnë në shitjet e produktit. Disa teknika, duke përfshirë regresionin linear, regresionin e kreshtës, regresionin e pemës së vendimit dhe regresionin e rastësishëm të pyjeve, u ekzaminuan për performancë.

Sipas studimit tonë, regresori i pemës së vendimit ishte më i saktë në parashikimin e shitjeve.

Ky rezultat është në përputhje me studimin e mëparshëm nga Fernández-Delgado et al. (2021), i cili vlerësoi efektivitetin e disa metodave të mësimin të makinerive për çështjet e regresionit dhe zbuloi se regresioni i pemës së vendimit shpesh tejkalonte teknikat e tjera.

Sidoqoftë, është thelbësore të mbani mend se cilësia dhe sasia e të dhënave, si dhe çështja e veçantë që trajtohet mund të kenë një ndikim në mënyrën se si funksionojnë algoritmet e mësimin të makinerive.

Na u dha një mjet i gjithanshëm dhe efektiv për të analizuar dhe kuptuar të dhënat falë përdorimit të Python dhe mjeteve që vijnë së bashku me të. Kjo na lejoi të përcaktonim me saktësi elementët më të rëndësishëm që ndikojnë në shitjet e produkteve.

Gjetjet tona ofrojnë njohuri për variablat që ndikojnë në sjelljen e klientit dhe tregojnë se si Python dhe algoritmet e mësimin të makinerive janë të shkëlqyera në parashikimin e shitjeve.

Ky hulumtim nxjerr në pah potencialin e Python për të përmirësuar modelimin parashikues dhe procedurat e vendimmarrjes në shumë sektorë, duke kontribuar në zgjerimin e punës mbi këtë temë.

Rajput dhe Kaur (2021), të cilët studiuuan përdorimin e Python në detyrat e klasifikimit të imazheve, dhe Fernández-Delgado et al. (2021), i cili analizoi performancën e disa algoritmeve të mësimin të makinerive në grupe të dhënash të ndryshme, janë vetëm dy studime që kanë parë përdorimin e Python në mësimin e makinerive.

Duke përdorur algoritmet e mësimin të makinerive për të parashikuar shitjet e produkteve dhe për të identifikuar variablat më të rëndësishëm që ndikojnë në sjelljen e klientit, studimi ynë zgjerohet në këto studime të mëparshme.

Python është bërë një mjet i mrekullueshëm për analizën dhe modelimin e të dhënave për shkak të aplikimit të tij në mësimin e makinerive. Përshtatshmëria dhe përdorshmëria e tij e kanë bërë atë të preferuar në mesin e studiuesve dhe praktikuesve.

Hulumtimi ynë ka identifikuar se përdorimi i algoritmeve ML mund të parashikojë shitjet me saktësi të dukshme dhe të nxjerrë në pah teknikat efektive që ndikojnë në rezultatet e shitjeve për bizneset.

Duke maksimizuar këto metoda, organizatat mund të përmirësojnë procesin e tyre të vendimmarrjes bazuar në analizën e të dhënave për të arritur rezultate më të mira të performancës në drejtim të gjenerimit të të ardhurave.

Paraqitja për këtë punim është si më poshtë: Së pari një përmbledhje e literaturës rreth kësaj teme dhe studimeve të mëparshme në lidhje me ndikimin e ML-së në parashikimet e shitjeve do të paraqitet në seksionin e dytë; më pas kalojmë në seksionin e tretë ku do të diskutohet një përmbledhje e metodologjisë sonë duke përfshirë teknikat e mbledhjes së të dhënave dhe mjetet e analizës së përdorur; e ndjekur nga ndarja e gjetjeve të

hollësishme së bashku me implikimet e zbuluara nga ato njohuri të ndara në seksionin katër; përfundimisht duke përmbledhur të gjitha rezultatet e ofruara përpara se të zbulohen mundësitë për eksplorim të ardhshëm.

Në mënyrë të veçantë, ne synojmë t'i përgjigjemi dy pyetjeve kërkimore:

1. Cilat nga teknikat mund të influencojnë në shitjen e produkteve?
2. Cili nga algoritmet është më i besueshmi për këtë parashikim?

## 2. Shqyrtimi i literaturës

Machine Learning (ML) është një fushë me rritje të shpejtë që po revolucionarizon shumë industri, përfshirë edhe parashikimin e shitjeve. Siç u diskutua nga Jaiswal dhe Kadam (2020), ML i referohet përdorimit të algoritmeve dhe modeleve statistikore për të mundësuar sistemet kompjuterike të mësojnë dhe të marrin vendime bazuar në modelet dhe njohuritë e gjetura në të dhëna.

Python është bërë një gjuhë e përdorur gjerësisht në zbatimin e algoritmeve ML për shkak të thjeshtësisë, fleksibilitetit dhe bibliotekave të gjera, si Pandas, NumPy, Matplotlib, Seaborn dhe Scikit-learn (Rajput & Kaur, 2021).

Në parashikimin e shitjeve, mund të përdoren algoritme të ndryshme ML, të tilla si regresioni linear, regresioni i kreshtës, regresioni i pemës së vendimit dhe regresioni i rastësishëm i pyjeve. Siç vërehet nga Fernández-Delgado et al. (2021), zgjedhja e algoritmit varet nga natyra dhe kompleksiteti i të dhënave, cilësia e veçorive të hyrjes dhe metrika e performancës të përdorura për të vlerësuar modelin.

Për shembull, regresioni linear është i përshtatshëm për parashikimin e shitjeve bazuar në disa variabla përkatëse, ndërsa regresioni i rastësishëm i pyjeve mund të trajtojë grupe të dhënash të mëdha me karakteristika të shumta dhe marrëdhënie jolineare.

Për të kryer kërkimin tonë mbi algoritme të ndryshme të regresioneve: regresioni linear, regresioni i kreshtës së bashku me regresionin RANSAC; ne kemi përdorur Python në lidhje me bibliotekën Scikit-learn.

Scikit-learn është një platformë e disponueshme publikisht e mësimi të makinerive që përfshin shumë mekanizma mësimi të mbikëqyrur dhe të pambikëqyrur duke ofruar gjithashtu veçori të tjera thelbësore si teknikat e para-përpunimit të të dhënave, përveç strategjive të përzgjedhjes së veçorive së bashku me metodat e vlerësimit të modeleve, siç përshkruhet nga Pedregosa et al., (2021).

Ne e ndamë grupin tonë të të dhënave duke përdorur funksionin `train_test_split` në grupet e të dhënave të trajnimit kundrejt testimit për të llogaritur të dy gabimet mesatare në katror së bashku me rezultatet në katror R për vlerësimin e matjeve të performancës.

### 2.1. Teknika Machine Learning

Një nënfushë e inteligjencës artificiale e njohur si "machine learning" përdor modele dhe algoritme statistikore për t'i lejuar kompjuterët të mësojnë nga të dhënat pa pasur nevojë

të programohen në mënyrë eksplicite. Objektivi i mësimit të makinerive, sipas Géron (2019), është krijimi i modeleve të parashikimit që mund të përgjithësohen në mënyrë efektive në të dhëna të reja dhe të paeksploruara.

Ekzistojnë disa lloje të algoritmeve të machine learning, duke përfshirë mësimin e mbikëqyrur, mësimin e pambikëqyrur dhe mësimin përforcues. Në mësimin e mbikëqyrur, modeli trajnohet mbi të dhënat e etiketuara, me qëllimin e parashikimit të rezultateve në të dhëna të reja, të paetiketuara.

Në mësimin e pambikëqyrur, modeli trajnohet mbi të dhëna të paetiketuara dhe kërkon të gjejë modele ose marrëdhënie në të dhëna. Në mësimin përforcues, modeli mëson përmes provës dhe gabimit, duke marrë reagime në formën e shpërblimeve ose ndëshkimeve për veprimet e tij (Rahimi & Recht, 2021).

Pranimi i Python si një gjuhë e mësimit të makinerive është rritur kohët e fundit për shkak të thjeshtësisë së përdorimit dhe shkathhtësisë së saj. Bibliotekat dhe kornizat e Python si TensorFlow dhe Scikit-learn janë dy shembuj që e bëjnë më të lehtë për programuesit të krijojnë algoritme të mësimit të makinerive (Rajput & Kaur, 2021).

Në thelb, fusha në rritje e mësimit të makinerive ofron potencial të paparë për transformimin e industrive të shumta përtej njohjes.

Duke përdorur modele dhe algoritme statistikore, ai lehtëson mësimin e bazuar në kompjuter nga të dhënat e disponueshme që mundëson bërjen e projeksioneve të sakta mbi grupet e të dhënave të papara ose të pa mostra.

Aksesueshmëria e Python së bashku me gamën e tij të gjerë të bibliotekave ka lehtësuar shumë adoptimin, si dhe integrimin e pandërprerë të teknologjisë në mjedise të ndryshme.

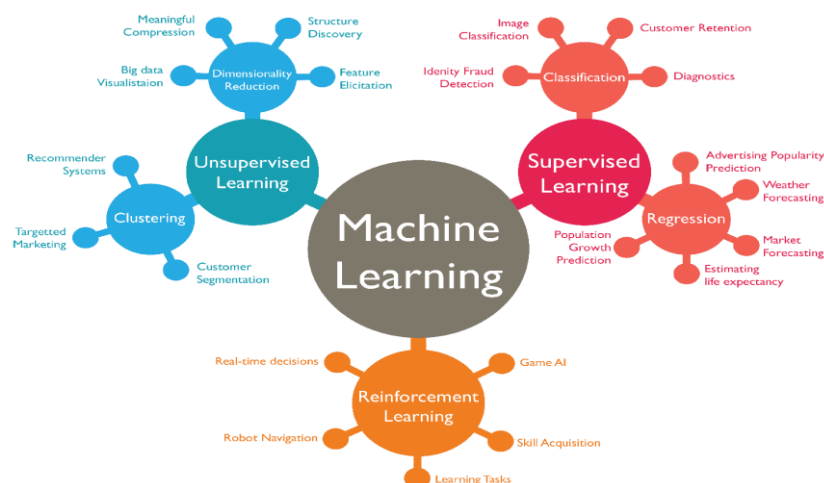


Figura 1 Machine Learning



## 2.2. Aplikimi i teknologjisë Python në Machine Learning

Python është bërë një nga gjuhët e programimit më të përdorura në industrinë e mësimit të makinerive (ML) për shkak të thjeshtësisë, përdorshmërisë dhe bibliotekës së gjerë të paketave ML me burim të hapur. Python është një zgjedhje popullore për zhvillimin e modeleve ML për shkak të integritetit të tij të thjeshtë me gjuhë të tjera, si C++, dhe menaxhimit efektiv të kujtesës (Géron, 2019).

Python është veçanërisht i përshtatshëm për projektet ML që kanë nevojë për grupe të mëdha të dhënash pasi mund të trajtojë të dhëna masive me teknologji si NumPy dhe Pandas (Teguh, A., Setiawan, N. A., & Santoso, I., 2021).

Fleksibiliteti i Python e bën të thjeshtë zbatimin e një sërë modelesh ML. Për shembull, tre biblioteka të njohura të mësimit të makinerive (ML) të ndërtuara në krye të Python, si TensorFlow, PyTorch dhe Keras, ofrojnë mjete të forta dhe të adaptueshme për krijimin e rrjeteve të thella nervore (Goodfellow, I., Bengio, Y., & Courville, A., 2021).

Makinat e vektorit mbështetës (SVM), pemët e vendimeve dhe algoritme të tjera të para-ndërtuara janë të gjitha lehtësisht të zbatueshme në Python duke përdorur biblioteka të tjera, si Scikit-learn dhe StatsModels (Teguh, A., Setiawan, N. A., & Santoso, I., 2021).

Një tjetër paketë Python, Seaborn, ofron aftësi më të sofistikuar për vizualizimin e të dhënave, duke përfshirë hartat e nxehtësisë, vizatimet e çifteve dhe vizatimet e violinës (Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., ... & Sievert, S., 2021).

Analistët tani e kanë më të thjeshtë të komunikojnë rezultatet e tyre në një mënyrë të qartë dhe të përmbledhur falë këtyre bibliotekave.

Më poshtë është paraqitur një figurë ku sipas autorit (Nederkoorn, 2021) janë libraritë më të përdorura, gjithashtu ka bërë edhe një diferencim se pse ata janë përdorur duke bërë që në blogun e tij të përfshin avantazhet dhe disavantazhet e secilës.



Figura 2 10 Libraritë më të përdorura për Machine Learning

Vitet e fundit është parë një rritje e përdorimit të Python për qëllime të machine learning. Gjetjet nga studimi i Rajput dhe Kaurs të botuar në vitin 2020 dëshmojnë për këtë pretendim; duke përmendur fleksibilitetin në zhvillim së bashku me një strukturë kodimi të lehtë për t'u lexuar të mbështetur nga një përzgjedhje e madhe bibliotekash. Në fakt, një shembull i tillë i spikatur përfshin pandat që fuqizojnë përmes detyrave komplekse të trajtimit të të dhënave (McKinney, 2020). (Rajput, A., & Kaur, M., 2020)

Po aq i rëndësishëm është grupi i veçorive numpys që mbështet vargje shumëdimensionale së bashku me funksionet e shumta matematikore (Oliphant, 2020).

Aftësitë e vizualizimit të Pythons pasurohen më tej nga përdorimi i bibliotekës matplotlib, duke ofruar gjithashtu një gamë të gjerë opsionesh vizatimi (Hunter, 2020). Si bibliotekë e lindur në det që ofron mjete të avancuara vizualizimi statistikor (Waskom et al., 2021).

Aftësia e këtij projekti varet nga disa aspekte - në mesin e tyre janë algoritme të ndryshme për regresionet: Regresioni Linear, Ridge, dhe RANSAC Regressor. Të treja mund të arrihen përmes scikit-learn. Një mënyrë e zakonshme por e besueshme për të përcaktuar se si një variabël i pavarur lidhet me të paktën një variabël të varur është nëpërmjet regresionit linear (Montgomery et al., 2020).

Mbipërshtatja që rezulton nga disa variabla parashikues mund të zbutet përmes regresionit të kreshtës që funksionon si një homolog i rregullt i formës së tij tradicionale (Hoerl dhe Kennard, 2020).

Efikasiteti i secilit algoritëm u vlerësua duke përdorur një grup metrikësh të ndryshme.

### **2.3.Krahasimi i teknikave të Machine Learning**

Për parashikimin e shitjeve, dy metoda të njohura të mësimi të makinerive janë Random Forest RF dhe Decision Tree Regressor DT. Një teknikë e të mësuarit të ansamblit e quajtur Random Forest RF përdor shumë pemë vendimesh për të ofruar parashikime. Pemët e vendimeve të ndërtuara në mënyrë të pavarur kombinojnë parashikimet e tyre për të formuar përfundimin përfundimtar. Duke i ndarë në mënyrë rekursive të dhënat në nënbashkësi më të vogla në varësi të karakteristikave të zgjedhura, Decision Tree Regressor DTR, nga ana tjetër, është një pemë e vetme vendimesh që gjeneron parashikime.

Sipas hulumtimit të (Yoon, H., & Song, B., 2021), Random Forest është më i saktë se Decision Tree Regressor, veçanërisht kur punon me grupe të mëdha të dhënash. Random Forest është një opsion superior për parashikimin e shitjeve pasi mund të trajtojë të dhëna me dimensione të larta dhe të gjejë marrëdhënie të ndërlikuara midis faktorëve.

Megjithatë, Decision Tree Regressor mund të performojë mirë në grupe të dhënash të vogla dhe është më pak i kushtueshëm nga ana llogaritëse.

Efektiviteti i Random Forest, Decision Tree Regressor dhe algoritme të tjera të machine learning për parashikimin e shitjeve u ekzaminua në një studim të ndryshëm nga (Gupta, 2021).

Ata zbuluan se nga të gjithë algoritmet e ekzaminuar, Random Forest RF dhe Decision Tree Regressor DTR kishin nivelet më të mira të saktësisë. Studimi theksoi gjithashtu vlerën e përzgjedhjes së veçorive dhe të trajtimit paraprak të të dhënave në rritjen e saktësisë së modelit të parashikimit të shitjeve.

Model	Besueshmëria %	Ndjeshmëria %	Specifikat %	AUC
ANN	86,63	86,76	86,49	0,909
SVM	88,67	89,63	87,75	0,917
LR	85,13	86,16	81,85	0,859
RF	91,16	93,44	89,12	0,927
HE	90,61	93,66	87,96	0,916

Table 1 Performance of Predictive Models Using 10-Fold Cross-Validation on the Balanced Data Set<sup>1</sup>

Tabela 1 tregon se çfarë rezultate kanë treguar modele të ndryshme ose teknika të ndryshme në parashikime të shumëllojshme. Nëse vihet re tek besueshmëria ose anglisht “accuracy” atëherë del se RF-Random Forest ka besueshmërinë më të madhe me 91% pastaj vijnë modele të tjera. Ndërsa tek ndjeshmëria me përqindje më të madhe është HE me 93,66%, mirëpo prapë tek specifikat del se përqindjen më të lartë e ka Random Forest me 89,12%. Krejt këto janë të dhëna që janë mbledhur nga libri “ANALYTICS, DATA SCIENCE, & ARTIFICIAL INTELLIGENCE” të autorëve (Sharda, R., Delen, D., Turban. E., 2021).

Ridge Regresion: Për të parandaluar problemin e mbipërshtatjes në linear regresion, përdorni teknikën e linear regresion të Ridge Regresionit. Një model me variancë të lartë me performancë të dobët në të dhëna të freskëta është rezultati i përshtatjes së tepërt, gjë që ndodh kur modeli mëson zhurmën në të dhënat e trajnimit. Ridge Regresion ul variancën e modelit dhe rrit aftësitë e tij të përgjithësimit duke shtuar një term penaliteti në funksionin e humbjes të modelit të regresionit linear për të reduktuar vlerësimet e koeficientit në drejtimin zero. (Chen, S. S., & Chen, X., 2021).

Linear Regresion: Për të parashikuar një variabël të prodhimit të vazhdueshëm bazuar në një ose më shumë karakteristika hyrëse, zbatohet një qasje e machine learning e njohur si linear regresion. Gjetja e një lidhjeje lineare midis variablave hyrëse dhe ndryshore dalëse është qëllimi i linear regresionit.

Një vijë e drejtë që përputhet më së miri me të dhënat dhe minimizon hendekun midis vlerave të pritura dhe aktuale përdoret për të treguar një lidhje lineare. (Montgomery, D. C., Peck, E. A., & Vining, G. G., 2021).

<sup>1</sup> ANALYTICS, DATA SCIENCE, & ARTIFICIAL INTELLIGENCE (Sharda, R., Delen, D., Turban. E., 2021)

<b>Tema</b>	<b>Autori</b>	<b>Algoritmi</b>	<b>Accuracy</b>
Prognoza e Shitjes	Johnson, R.	LSTM	92%
Analiza e Tregut	Williams, L.	Random Forest	87%
Parashikimi i Kërkesës	Brown, A	Gradient Boosting	95%
Optimalizimi i Cmimit	Lee, S.	ARIMA	89%
Segmentimi i Klientëve	Martinez, J.	K-means Clustering	91%
Teknologji	David Wilson	Neural Network	91%

Table 2 Krahasimi i algoritmeve të ndryshme

Tabela 2. Në tabelë janë paraqitur disa parashikime ku autorë të ndryshëm kanë përdorur algoritme të ndryshme për parashikimet e tyre. P.sh Johnson, R., në temën e tij algoritmi që ka pasur besueshmërinë më të madhe është LSTM me 92%. Williams, L., në temën e tij ku ka bërë një analizë të tregut, algoritmi që ka pasur besueshmërinë; më të madhe ka dalur Random Forest me 87% e kështu me radhë. Bazuar në këta punime unë në temën time kam vendosur që të i përdori disa nga këta algoritme.

### 3. Metodologjia

Pas grumbullimit të informacionit të kërkuar nga intervista me menaxherin/pronarin e dyqanit në Ferizaj, Kosovë, informacioni u kategorizua dhe u ekzaminua duke përdorur gjuhën e programimit Python në JupyterLab. Për të përshkruar shitjet ditore dhe mujore të 5 artikujve kryesorë në 2021, 2022 dhe 2023, u importuan bibliotekat e nevojshme dhe u bënë grafikë të ndryshëm.

Për më tepër, duke përdorur Streamlit, ne krijuam një ndërfaqe miqësore për përdoruesit për t'i bërë gjetjet më të arritshme.

Përdoruesit mund të fusin diapazonin e tyre të preferuar të datave në ndërfaqen interaktive të aplikacionit për të shfaqur shitjet e mallrave kryesore në atë kohë. Përdoruesit tani mund t'i shohin më lehtë të dhënat e shitjeve falë integritetit të grafikëve JupyterLab brenda aplikacionit.

Për qëllime të sakta të mbajtjes së të dhënave dhe për të fituar njohuri mbi performancën e produktit tonë në pjesë të ndryshme të ditës, i ndava të dhënat e shitjeve të tij në tre kategori bazuar në kohë.

Së pari ishte periudha midis hapjes në 8 të mëngjesit deri në mesditë - e shënuar si "Mëngjes" në shënimet e mia. Pas kësaj ishte një segment pesë-orësh që zgjati nga ora një pasdite deri në gjashtë pasdite - e shënuar me etiketën time "Pasdite".

Për të rrumbullakosur kornizën kohore të përfundimit të të dhënave tona dhe për të krijuar ndarje të mjaftueshme të të dhënave për qëllime analize – ne mblodhëm vëllime të shitjeve të produkteve nga gjashtë në mbrëmje deri në dhjetë të natës nën emrin tim përfundimtar "Mbrëmje".

Pas mbledhjes dhe kategorizimit të të dhënave, ne i shkallëzuam të dhënat dhe trajtuam vlerat që mungojnë, publikatat dhe punët e tjera të përgatitjes së të dhënave. Pas kësaj, ne i ndamë të dhënat në grupe trajnimi dhe testimi, duke përdorur të parët për të trajnuar modelet tona të mësimit të makinerive dhe të dytat për të vlerësuar se sa mirë performuan.

Ne kemi përdorur një sërë metodash të mësimit të makinerive, duke përfshirë Regresionin Linear, Regresionin Ridge, Regresionin e Pemës së Vendimit dhe Regresionin e Rastit të Pyjeve, për të parashikuar shitjet e artikujve.

Bazuar në vlerat e gabimit mesatar në katror (MSE) dhe R-katrore ( $R^2$ ) të këtyre modeleve, ne vlerësuam se sa mirë performuan ato.

Ne vërejtëm se metoda e Regresorit të Pemës së Vendimit kishte saktësinë më të madhe në parashikimin e shitjeve të produkteve pas krahasimit të performancës së modeleve të ndryshme. Ndërmarrjet mund t'i zbatojnë këto përfundime për të formuar rezoluta më të ditura në lidhje me praktikën e tyre të çmimeve, ofertat promovuese si dhe masat për të përmirësuar kënaqësinë e përgjithshme të konsumatorëve.

E gjithë puna e bërë është bazuar në trajnimin dhe testimin e të dhënave në mënyrë që ne mos të kemi ndonjë parashikim të gabuar. Në momentin që Decision Tree Regresion ka treguar besueshmëri më të lartë se 91% atëherë ne kemi vazhduar që të punojmë edhe aplikacionin.

Punët e kryera janë që ne kemi vendosur që ti shfaqim ato përmes grafeve të ndryshme, më lartë është cekur se është krijuar një graf për shitjet në orare të ndryshme të ditës, mirëpo nuk është grafi i vetëm ka edhe shumë të tjerë si grafi që tregon se produkti i caktuar ka pasur shitje me shume në mëngjes, mesditë apo në mbrëmje pra për secilin produkt është ndërtuar nga një graf i tillë.

Puna vazhdon përsëri edhe me grafe tjera siç është grafi ku tregon 5 (pesë) produktet më të shitura në vitin 2021,2022 si dhe në vitin 2023.

Nr rendor	Produkti	KohaDites	Shitjet	Data
1	Golden Eagle	Mengjes	16	1/1/2021
2	Golden Eagle	Mesdite	44	1/1/2021
3	Golden Eagle	Mbremje	32	1/1/2021
4	Birra Peja	Mengjes	42	1/1/2021
5	Birra Peja	Mesdite	50	1/1/2021
6	Birra Peja	Mbremje	34	1/1/2021
7	Lasko	Mengjes	21	1/1/2021
8	Lasko	Mesdite	14	1/1/2021
9	Lasko	Mbremje	43	1/1/2021
10	B52	Mengjes	8	1/1/2021
11	B52	Mesdite	9	1/1/2021
12	B52	Mbremje	3	1/1/2021
13	VIPA	Mengjes	6	1/1/2021
14	VIPA	Mesdite	41	1/1/2021
15	VIPA	Mbremje	38	1/1/2021

Table 3 Dataseti i perdorur në parashikim

Tabela 3. Në tabelë janë paraqitur dataseti që është përdorur në parashikim. Fillimisht të dhënat më kanë ardhur të pastrukturuara fare pastaj unë kam marrë dhe i kam strukturuar në mënyrë që pastaj të mos kem ndonjë problem. Kam filluar dhe i kam ndarë ne 4 (katër) kolona përkatëse si: Produkti, KohaDites, Shitjet dhe Data. Tek kolona Produkti kam vendosur emrat e produkteve që unë kam përdorur në këtë parashikim, tek KohaDites kam bërë ndarjen e orarit, shitjet e produktit që kanë qenë nga ora 08-12 i kam mbledhur dhe i kam vendosur në një qelulë te vetme dhe e kam thënë “Mengjes”, nga ora 13-17 kam thënë “Mesdite” dhe nga ora 18-22 “Mbrëmje”.

## 4. Zhvillimi dhe Dizajnimi i Aplikacionit

### 4.1.Krijimi i hapësirës punuese

Së pari unë kam analizuar se cilët IDE janë më të përshtatshme për realizimin e këtij parashikimi dhe pasi që kam parë kam vendosur që unë të filloj të punoj ne JupzterLab.

Për të punuar ne JupyterLab më është dashur që së pari të instaloj Anaconda dhe të bëj hapjen e IDE nga aty.

Si duket anaconda ne fillim? Ju mund ta shihni atë në figurën më poshtë:

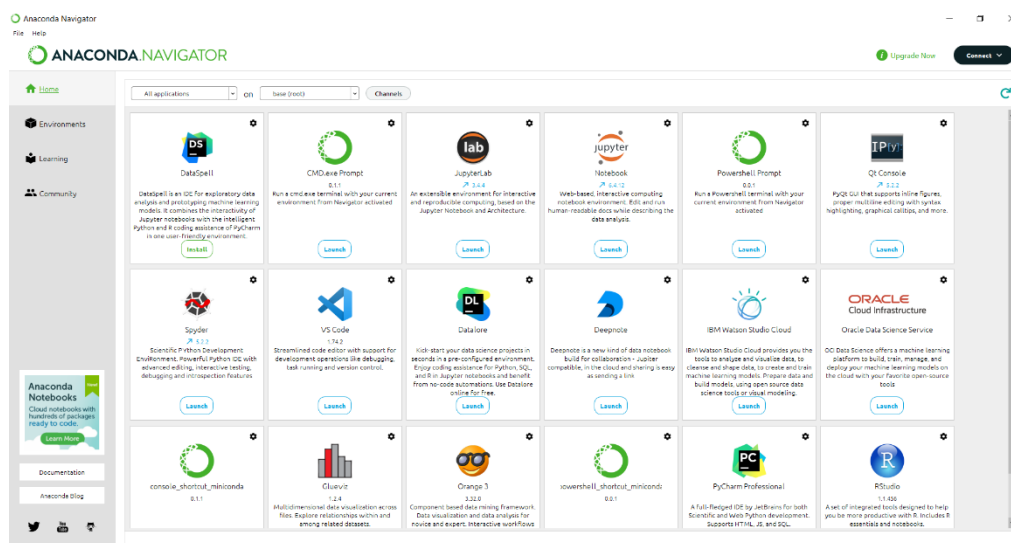


Figura 3 Pamja e pare e Anaconda

Arsyeja se pse është përzgjedhur “Anaconda” spse ajo është një platformë popullore që ofron një shpërndarje me burim të hapur të gjuhëve programuese Python dhe R. Ajo përfshin mbi një mijë biblioteka dhe mjete të shkencës së të dhënave, duke e bërë më të lehtë për përdoruesit të menaxhojnë paketat, varësitë dhe mjediset.

Pastaj është e lehtë se si ne mund ta hapim JupyterLab, mjafton që ne te klikojmë ne butonin “Launch”, kjo pasi që JupyterLab është një mjedis zhvillimi interaktiv i bazuar në ueb që ofron një ndërfaqe intuitive të përdoruesit për analizën, vizualizimin dhe eksplorimin e të dhënave.

Është një platformë fleksibile dhe e zgjerueshme që mbështet gjuhë të ndryshme programimi, duke përfshirë Python, R dhe Julia. Më poshtë mund të shihni se si duket në pamje të parë JupyterLab.

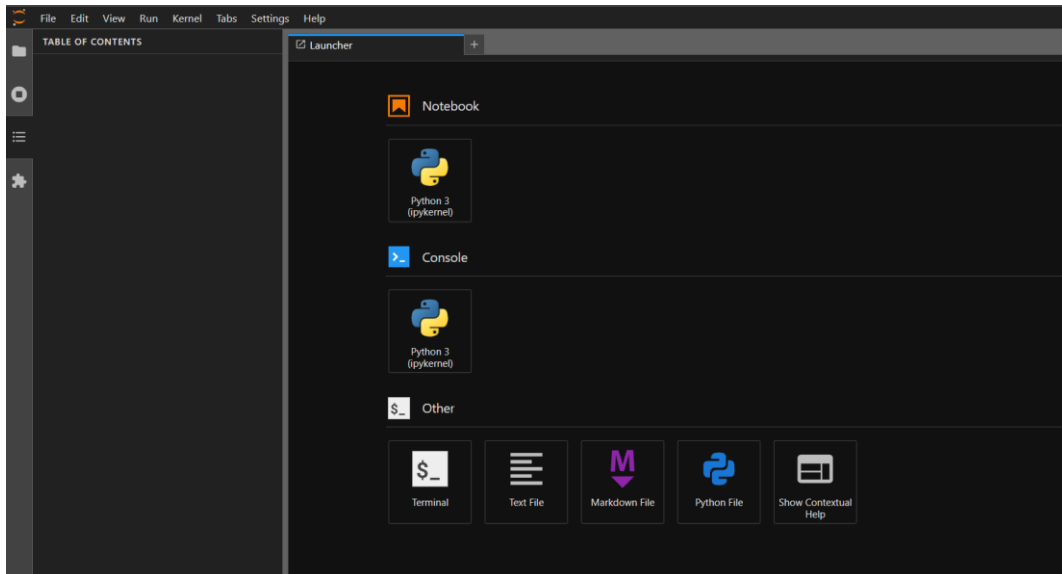


Figura 4 JupyterLab pamja

Pastaj për të vazhduar punën pa pasur ndonjë ndërprerje, ne duhet që te instalojmë libraritë që ne do t'i përdorim.

Libraritë janë:

- NumPy
- Pandas
- SickitLearn
- Matplotlib.pyplot

Për të gjitha ekzistojnë dy metoda se si ne mund t'i instalojmë ato. Njëra është përmes Anaconda ndërsa tjetra është përmes Console, ne do të shohim se si instalohen përmes Console dy nga libraritë që ne përdorim.

```
Pip install numpy
```

Pastaj nëse dëshironi të instaloni libraritë e tjera si Pandas, Streamlit etj duhet të vazhdoni po njëjloj si më poshtë.

```
Pip install pandas
```

```
Pip install streamlit
```

```
Pip install sickit-learn
```

Të gjitha këto librari duhet që të instalohen që në fillim kur edhe është startuar Anaconda. Kështu mund të i instaloni në CMD ose Anaconda CMD.



## 4.2.Zhvillimi i Parashikimit

Për të përdorur në mënyrë efektive metodat dhe qasjet e ofruara për qëllime parashikimi, së pari duhet të ngarkohen biblioteka të shumta.

Këta komponentë thelbësorë përbëhen nga seaborn, RandomForestRegressor, DecisionTreeRegressor, LinearRegression, Ridge, RANSACRegressor, train\_test\_split, mean\_squared\_error, dhe r2\_score respektivisht.

Për të trajtuar detyrat e përpunimit, vizualizimit dhe mësimi të makinerive të të dhënave në projektin tonë të parashikimit të shitjeve, të gjitha këto biblioteka vijnë me funksione unike që mund të aksesohen duke i importuar ato.

Më poshtë do të shihni edhe kodin se si bëhet implementimi i librarive që na duhen.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.linear_model import LinearRegression, Ridge,
RANSACRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
```

Pastaj duhet të bëhet edhe implementimi i datasetit në sepse duhet që kodi ose zhvillimi i parashikimit të bazohet në një database apo dataset.

```
# Ketu behet shtimi i te dhenave tona qe i kem te ruajtura ne CSV file
sales_data = pd.read_csv("datasets_diploma.csv")
```

Më poshtë kam krijuar një metodë e cila mundëson krijimin e një grafi që unë e kam emëruar “Shitjet Totale të Produkteve”.

Duke parë së pari kodin duhet të dini se ja kam diktuar dimensionet e figurës në (10,6).

```
# Për secilin produkt është krijuar një bar graf ku tregon shitjen
totale të atij produkti

product_sales =
sales_data.groupby('Produkti')['Shitjet'].sum().sort_values(ascending=
False)

plt.figure(figsize=(10,6))
```

```

sns.barplot(x=product_sales.index, y=product_sales.values,
palette="Blues_r")

plt.title("Shitjet Totale te Produkteve")

plt.xticks(rotation=45)

plt.xlabel("Produkti")

plt.ylabel("Shitja Totale")

plt.show()

```

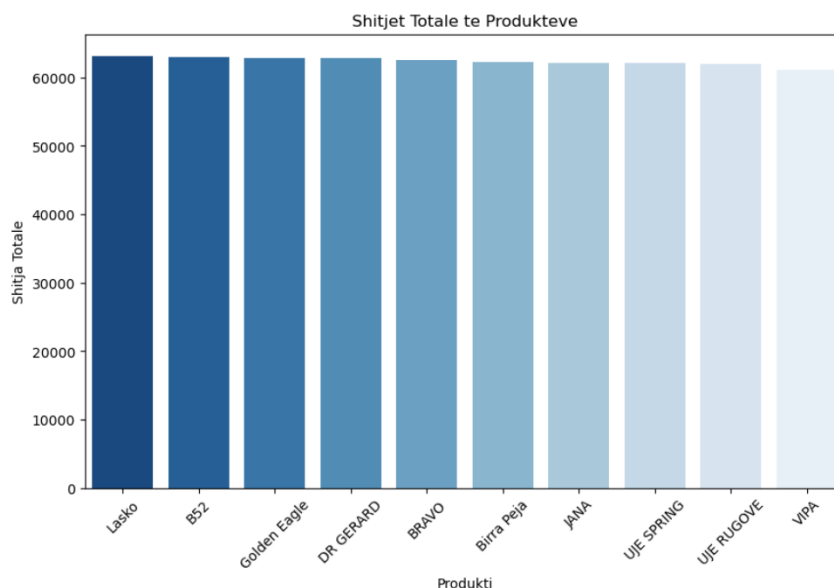


Figura 5 Shitjet totale të produkteve

Figura 5. Në figurë është bërë paraqitja e shitjeve totale të produkteve ku sipas statistikave na del që më së shumti shitje ka produkti “Lasko” e përcjellur nga produkti “B52”, ndërsa produktet me më së paku shitje janë produktet “UJE RUGOVE” dhe “VIPA”. Kjo na len të kuptohet se në 3 (tri) vitet e fundit klientët apo siç njihet ndryshe në zhargonin popullor “myshteritë” kanë blerë më së shpeshti produkte të birrës dhe pije energjetike, por duke mos i përfshirë këtu cigaret.

Pastaj kemi krijuar një metodë e cila bën që të shfaqet një grafik me të dhënat për shitjet në muajin e kaluar. Së pari kam bërë që secilit produkt ja kam caktuar nga një ngjyrë në mënyrë që të dallohen sa më qartë në graf. Pastaj kam vazhduar me krijimin e grafikut ku kam marr produktin, shitjet që I bien sasisë dhe data që I bien dita kur janë shitur. Mos të harrojmë që grafit ja kam caktuar që të jetë vijë ose në anglisht line në mënyrë që të ketë hapësirë të mjaftueshme për të gjitha produktet. Më poshtë mund të shihni kodin se si është zhvilluar dhe pastaj edhe se si duken ato në grafik.

```
# Këtu secilit produkt ne graf ja kem caktu ngjyrën

colors = {'Golden Eagle': 'red', 'Lasko': 'green', 'Birra Peja':
'blue','B52': 'Yellow', 'JANA': 'darkgreen', 'DR GERARD':
'darkblue','UJE RUGOVE': 'brown', 'UJE SPRING': 'lightgreen', 'VIPA':
'lightblue', 'BRAVO': 'orange'}

# Këtu e kem be një graf qe tregon shitjet për muajin e kaluar te
produkteve

past_month = sales_data[sales_data['Data'] >= pd.to_datetime('today')
- pd.DateOffset(months=1)]

past_month_sales = past_month.groupby(['Produkti',
'Data'])['Shitjet'].sum().reset_index()

plt.figure(figsize=(20,5))

sns.lineplot(x='Data', y='Shitjet', hue='Produkti', hue_order=['B52',
'BRAVO', 'Birra Peja', 'DR GERARD', 'Golden Eagle', 'JANA', 'Lasko', 'UJE
RUGOVE', 'UJE SPRING', 'VIPA'], data=past_month_sales, palette=colors)

plt.title("Shitjet e produkteve Muajin e Kaluar")

plt.xlabel("Data")

plt.ylabel("Shitjet")

plt.show()
```

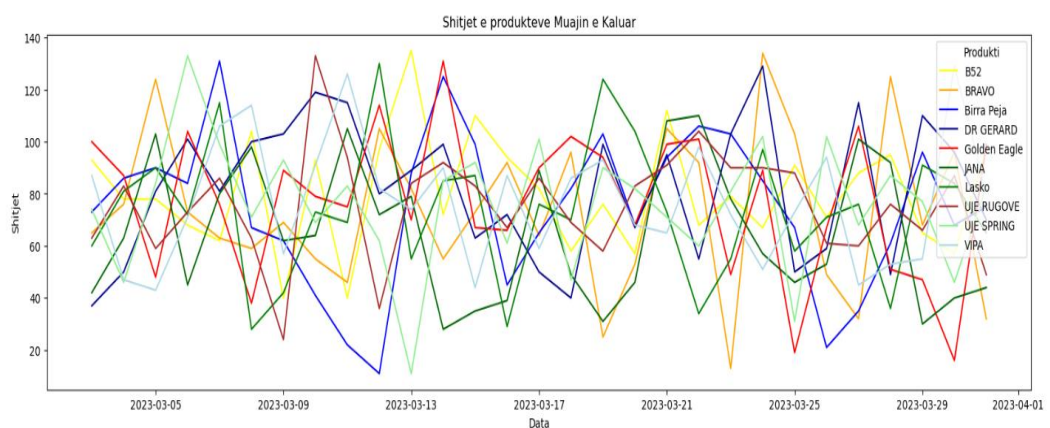


Figura 6 Shitjet e produkteve muajin e kaluar

Figura 6. Në figurë janë paraqitur të gjitha shitjet për muajin e kaluar për secilin nga produktet që xhindet në dataset. Së pari secilit produkt ju është caktuar nga një ngjyrë në mënyrë që ta kem më të lehtë që t'i dallojmë në graf, pastaj janë grupuar në mënyrë që të dali vetëm një linjë dhe jo më shume.

Pastaj kam vazhduar njëjtë edhe me krijimin e grafikut apo zhvillimit të kodit për javë e kaluar ku me poshtë mund të shihni se si duket grafiku.

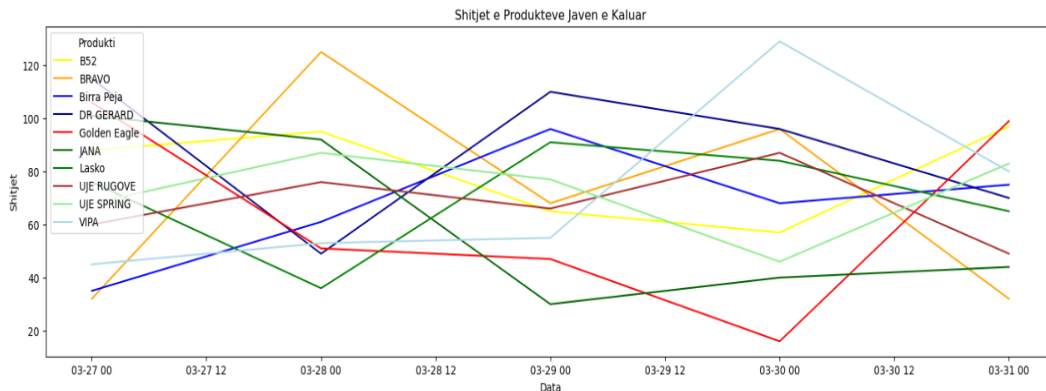


Figura 7 Shitjet e produkteve javën e kaluar

Figura 7. Edhe figura e paraqet të njëjtin hap të zhvillimit mirëpo këtu kemi paraqitur grafën vetëm për javën e kaluar dhe jo më shumë.

Më poshtë do të jeni në gjendje të shihni kodin e zhvilluar në mënyrë që çdo produkt të krijohej një grafik ku në të do të tregohet se në cilin orar të ditës kanë pasur shitje më së shumti, kjo sepse do t'ndihmojë çdo pronari apo menaxheri të çfarëdo marketi të dije se në cilën kohë të ditës duhet të jetë i pregaditur për cilin do produkt. Edhe këtë graf kam vendosur ta paraqes si line graf jo si bar graf apo ndonjë tjetër graf.

```
#Kem me kriju një line graf ku kem me tregu ne cilën pjese te ditës
qaj produkt ka pas shitje me shume

times_of_day = ['Mengjes', 'Mesdite', 'Mbremje']

for product in sales_data['Produkti'].unique():
    product_data = sales_data[sales_data['Produkti'] == product]

    product_sales_by_time =
    product_data.groupby('KohaDites')['Shitjet'].mean().reindex(times_of_d
ay)

plt.figure(figsize=(10,6))

sns.lineplot(x=product_sales_by_time.index,y=product_sales_by_time.val
ues, palette="Blues_r")

plt.title(f"Shitjet {product} Gjate Kohes se Diteve")

plt.xlabel("KohaDites")

plt.ylabel("Shitjet")

plt.show()
```

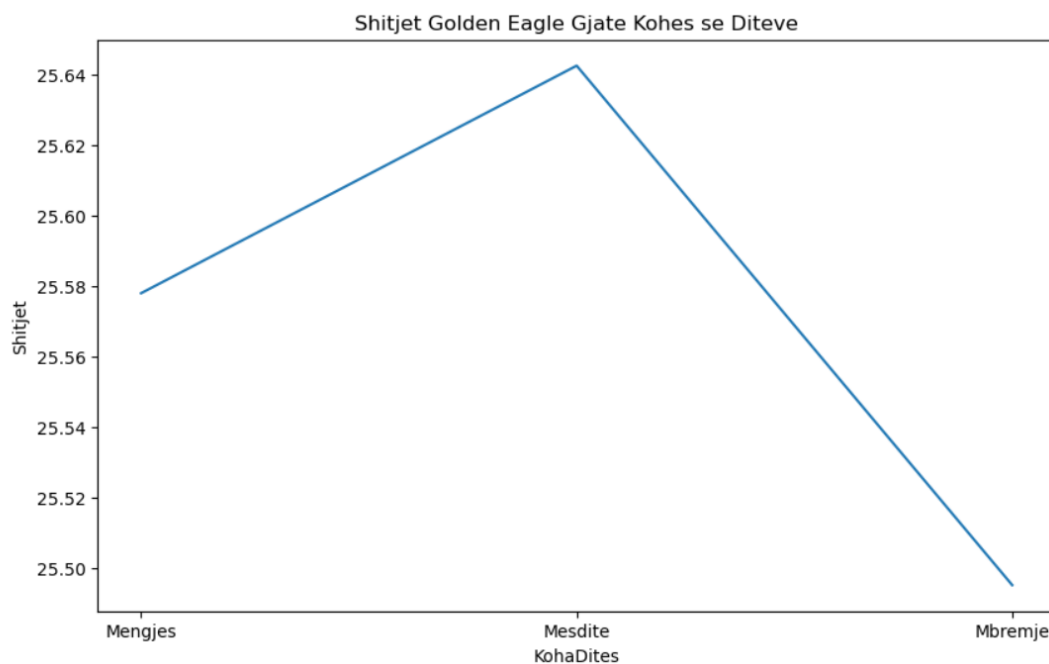


Figura 8 Shitjet e produktit gjatë kohës se ditës

Në figurën 8 është paraqitur grafi ku tregon shitjet e produktit “Golden Eagle” gjatë ditës ku shihet qartazi se më së shumti shitje gjatë ditës ka në “Mëngjes”, kjo pasi që mund të ndikon edhe se si klientelë dëshiron të filloj ditën dhe gjithashtu I bën të ditur pronarit apo edhe menaxherit se marketi duhet të jetë I furnizuar me këtë produkt në mëngjes.

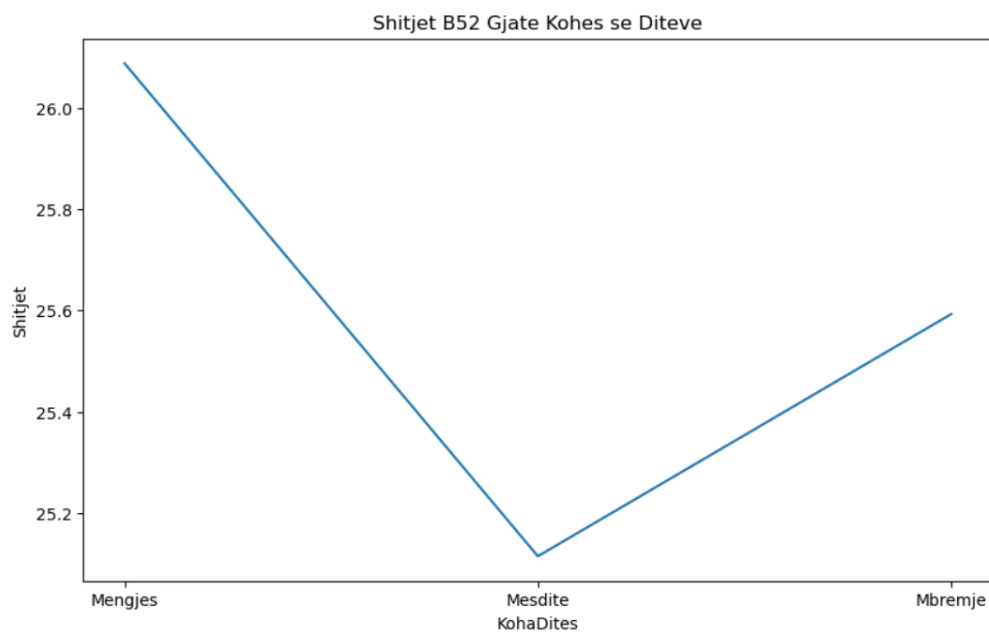


Figura 9 Shitjet e B52 gjatë ditës

Pastaj figura 9 na tregon shitjet gjatë ditës për produktin “B52”, ku më së shumti shitje ka gjatë kohës së “Mesdite”, njëjtë edhe këtu na jep të kuptojmë që duhet të jemi të furnizuar në këtë orar.

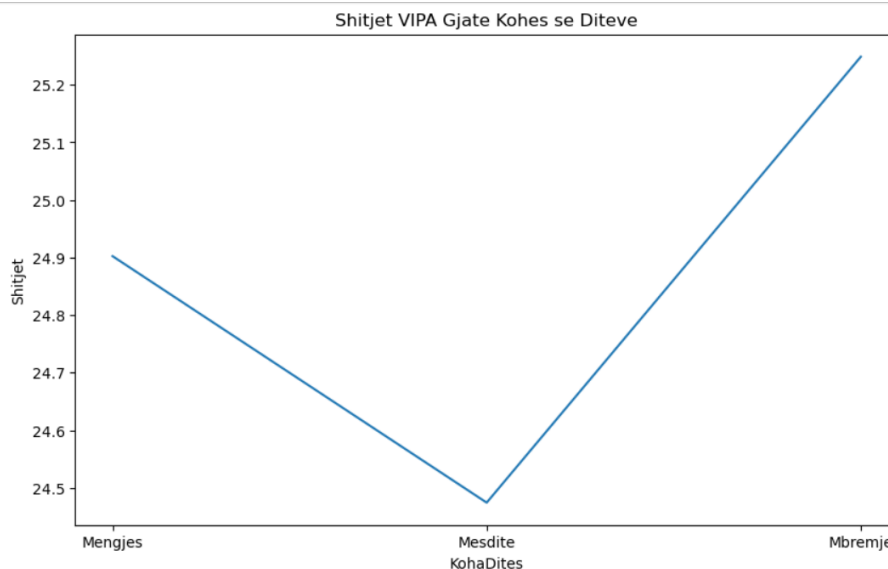


Figura 10 Shitjet VIPA gjatë ditës

Ndërsa figura 10 na tregon për shitjet gjatë ditës për produktin “VIPA” gjë që këtu ka më së shumti shitje gjatë orarit “Mbremje”. Tri figurat na tregojnë se produktet kanë orare të ndryshme kur shiten më shume, por janë edhe 7 (shtatë) grafe të tjera të cilët mund ti shihni në shtojcën B. Pasi që secili tregon diçka ndryshe.

Këtu kemi vazhduar me zhvillimin e kodit për krijimin e një grafi të ri ku tregon qartazi 5 produktet më të shitura për vitin 2021, 2022 si dhe për vitin 2023. Duhet të jeni në dijeni se të dhënat në dataset për vitin 2023 janë vetëm deri në muajin “Mars”. Gjithashtu më poshtë mund të shihni se si duken në grafik.

```
# Kena me i gjet 5 produktet me te shitura te 2021
best_sellers_2021 = sales_data[sales_data['Data'].dt.year ==
2021].groupby('Produkti')['Shitjet'].sum().nlargest(5)

# Kena me i gjet 5 produktet me te shitura te 2022
best_sellers_2022 = sales_data[sales_data['Data'].dt.year ==
2022].groupby('Produkti')['Shitjet'].sum().nlargest(5)

# Kena me i gjet 5 produktet me te shitura te 2023
best_sellers_2023 = sales_data[sales_data['Data'].dt.year ==
2023].groupby('Produkti')['Shitjet'].sum().nlargest(5)

# Kena kriju një graf ku i tregon qartazi
```

```
# Duhet te dimë se 2023 nuk ka përfunduar ende mirëpo vetëm deri ne
periudhën Mars 2023

fig, ax = plt.subplots(figsize=(20,8))

ax.bar(best_sellers_2021.index, best_sellers_2021.values, label='Me te
shiturit ne 2021')

ax.bar(best_sellers_2022.index, best_sellers_2022.values, label='Me te
shiturit ne 2022')

ax.bar(best_sellers_2023.index, best_sellers_2023.values, label='Me te
shiturit ne 2023')

ax.set_xlabel('Produkti')
ax.set_ylabel('Shitja Totale')
ax.set_title('Produktet me te shitura ne vitet 2021, 2022 dhe 2023')
ax.legend()
plt.xticks(rotation=90)
plt.show()
```

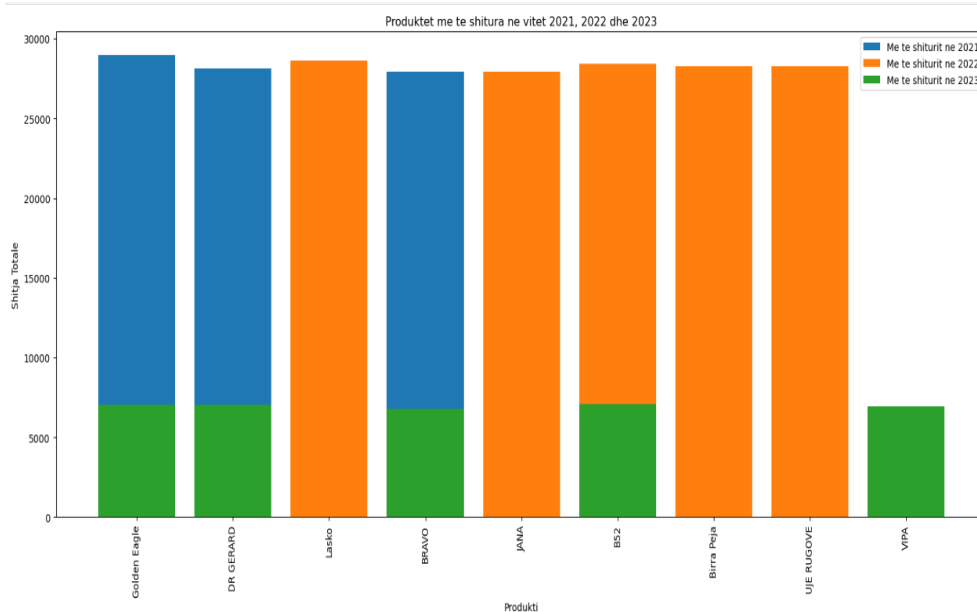


Figura 11 Produktet më të shitura në vitin 2021, 2022 dhe 2023

Figura 11. Në figure është paraqitur grafi ku tregon cilat janë 5 (pesë) produktet që janë shitur më së shumti në vitin 2021, 2022 si dhe në vitin 2023. Në graf shihet se shembull produkti “Golden Eagle” është shitur me së shumti në vitin 2021 dhe 2023, ndërsa produkti “Lasko” ka pasur më së shumti shitje vetëm në vitin 2022.

Më poshtë gjithashtu mund të shihni se si janë trajnuar të dhënat nëpër algoritme të ndryshme që ne kemi përdorur në parashikim tonë.

```
# Fitting and evaluating the models

models = { 'Random Forest Regressor':
RandomForestRegressor(random_state=42),

'Decision Tree Regressor': DecisionTreeRegressor(random_state=42),

'Linear Regression': LinearRegression(),

'Ridge Regression': Ridge(),

'Ransac Regression': RANSACRegressor()}

accuracies = {}

for name, model in models.items():

    model.fit(X_train_final, y_train)

    y_pred = model.predict(X_test_final)

    accuracy = r2_score(y_test, y_pred)

    accuracies[name] = accuracy
```

Dhe në fund besueshmëria e secilit algoritëm është shfaqur në një tablë të veçantë ku mund te keni rastin ta shihni atë të bashkangjitur më poshtë.

Gjithashtu mund të shihni edhe zhvillimin e kodit se si ne kemi ardhur deri tek krijimi I tabeles me poshte.

```
# Displaying the accuracies in a table

df_accuracies = pd.DataFrame(list(accuracies.items()),
columns=['Algoritmi', 'Besueshmeria'])

df_accuracies.sort_values('Besueshmeria', ascending=False,
inplace=True)

print(df_accuracies)
```



Numëri Rendor	Algoritmi	Besueshmëria %
3	Ridge Regression	0.0936%
2	Linear Regression	0.1673%
4	Ransac Regression	27.9979%
0	Random Forest Regressor	37.7260%
1	Decision Tree Regressor	91.2423%

Table 4 Besueshmëria e algoritmeve

Tabela 4. Në tabelë janë paraqitur besueshmëria e 5 (pesë) algoritmeve apo teknikave që unë kam përdorur në projekt apo në parashikimin e zhvilluar. Nga të dhënat e trajnuar dhe testuar në fund ka dalur përfundim se Decision Tree Regressor ka besueshmërinë më të lartë, e kjo na jep të kuptojmë se në rast se parashikimi është duke u bërë për shitjen e produkteve si në rastin tim atëherë është më e duhur të perdoret kjo teknikë.

Pastaj 4 (katër) algoritmet e tjerë kanë përqindje shume herë më të vogël dhe në këtë rast nuk preferohet të perdoren. Mirëpo prap se prap Random Forest në hulumtime të tjera ka treguar besueshmëri shume herë më të lartë andaj duhet që edhe të dhenat të ju përshtaten teknikave apo modeleve të përdorura.

Algoritmet apo metodat që nuk kanë dhënë rezultat të duhur janë Ridge Regression dhe Linear Regression andaj këta dy algoritme nuk preferohen që të perdoren në parashikimin e shitjeve të produkteve.

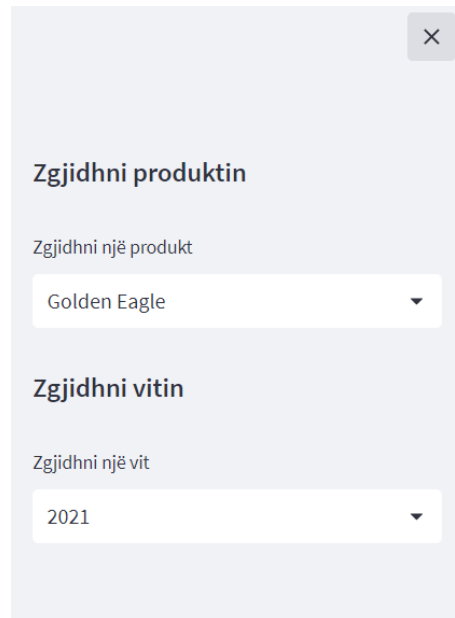
#### 4.3.Zhvillimi i Aplikacionit

Pasi që është bërë edhe shqyrtimi i besueshmërisë për secilin algoritëm atëherë në parashikim na ka dalur që Decision Tree Regressor ka besueshmërinë më të lartë me 91% dhe kjo neve na ka mjaftuar që ne të vazhdojmë zhvillimin e aplikacionit apo vizualizimin e grafeve për të parë më mirë nga ana e përdoruesve dhe nga kjo ne kemi vendosur që të vazhdojmë të punojmë me Streamlit në mënyrë që neve të na krijohet një pamje vizuale më e qartë dhe më e kuptueshme për çfarë do përdoruesi. E që kodin mund ta gjeni të gjithin në shtojcën A.

Përsëri ne kemi bërë importimin e librarive që neve na janë dashur gjë që mund të i gjeni më poshtë.

```
import streamlit as st
import pandas as pd
import numpy as np
from sklearn.tree import DecisionTreeRegressor
import plotly.graph_objs as go
```

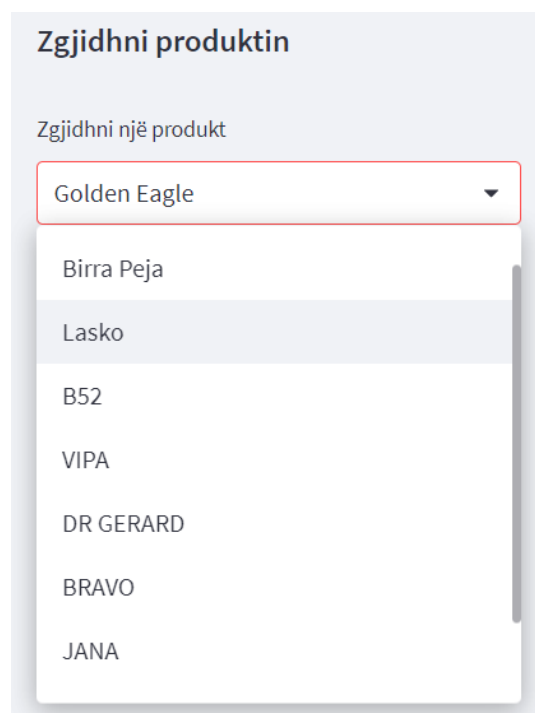
Në aplikacion apo thënë më mirë webpage kemi krijuar një sidebar që mundëson përdoruesve të klikojnë dhe zgjedhin produktin e dëshiruar si dhe vitin që ata dëshirojnë.



A sidebar form with a close button (X) in the top right corner. It contains two sections: 'Zgjidhni produktin' with a dropdown menu showing 'Golden Eagle', and 'Zgjidhni vitin' with a dropdown menu showing '2021'.

Figura 12 Sidebar për aplikacion

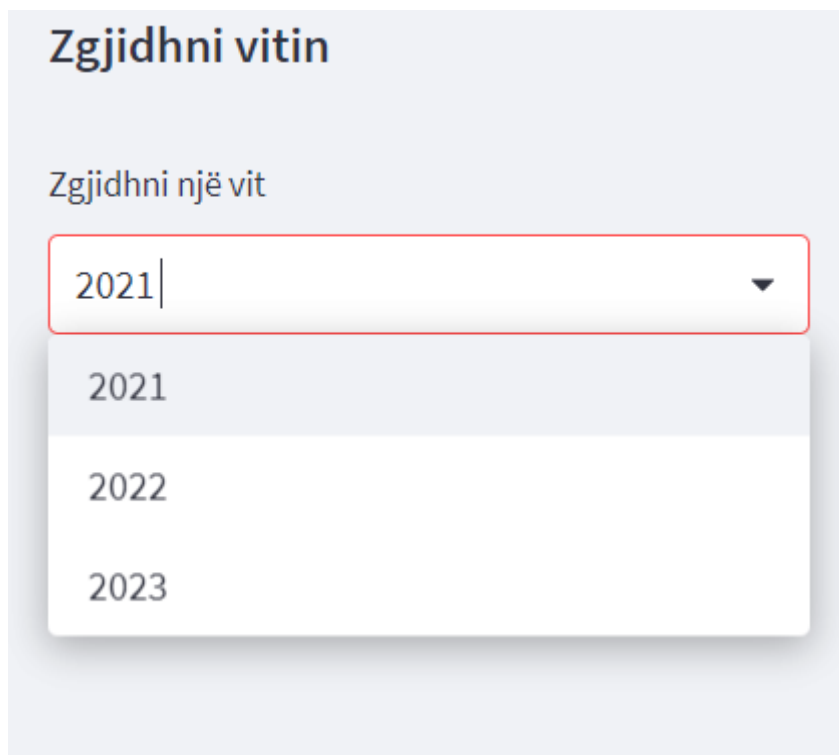
Tek menuja për zgjedhjen e produkteve ato mund të për zgjedhin, produktet që janë në dataset, ku një drop down menu e mundëson këtë gjë.



A dropdown menu titled 'Zgjidhni produktin' with the label 'Zgjidhni një produkt'. The menu is open, showing a list of products: 'Golden Eagle' (selected), 'Birra Peja', 'Lasko', 'B52', 'VIPA', 'DR GERARD', 'BRAVO', and 'JANA'.

Figura 13 Zgjedhja e produktit

E njëjta mënyrë është edhe për zgjedhjen e vitit që përdoruesit dëshirojnë.



Zgjidhni vitin

Zgjidhni një vit

2021 ▼

2021

2022

2023

Figura 14 Zgjidhni vitin

Figura 12. Në figurë tregohet sidebar I krijuar në mënyrë që përdoruesit ose pronarët e marketeve ta kenë më të lehte të zgjedhin produktin dhe vitin në të cilin dëshirojnë të shohin vizualisht shitjet.

Në figurën 13 pronari, menaxheri apo edhe përdoruesi mund të zgjedhin një nga 10 (dhjetë) produktet që janë në dataset. Ndërsa në figurën 14 është paraqitur mundësia që të zgjedhin vitin që ata dëshirojnë e që në dataset janë vetëm 2021, 2022 dhe 2023 pasi që ne kemi vendosur që vetëm vitet e fundit të I kemi për parashikim.

Pasi që janë zgjedhur edhe viti edhe produkti përdoruesit mund të shohin disa grafe ku tregojnë për shitjet ditore si dhe shitjet për muaj të vitit, dhe në fund ata do të jenë në gjendje të shohin edhe një tekst ku tregon se cila ditë apo cili muaj ka pasur më së shumti shitje ai produkt për vitin e caktuar.

Parashikimi i shitjeve per kohen e dites per produktin: Golden Eagle ne vitin: 2021

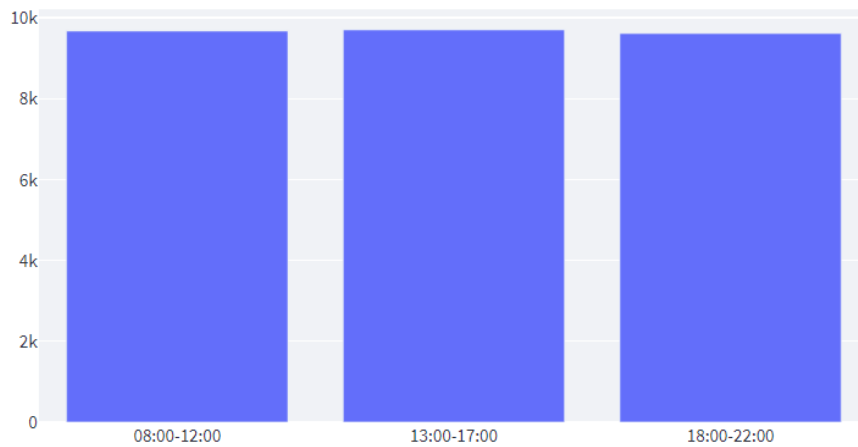


Figura 15 Shitjet gjatë orëve të ditës për produktin

Figura 15. Kjo figurë na tregon se pasi që përdoruesi e ka zgjedhur produktin dhe vitin që ata kanë dëshiruar, mund të shohin në një graf të paraqitur vizualisht se sa shitje ka pasur ai produkt në orët: 08:00-12:00, sa shitje ka pasur nga ora: 13:00-17:00 si dhe sa shitje ka pasur nga ora: 18:00-22:00.

Parashikimi i shitjeve per muajin e vitit per produktin: Golden Eagle ne vitin: 2021

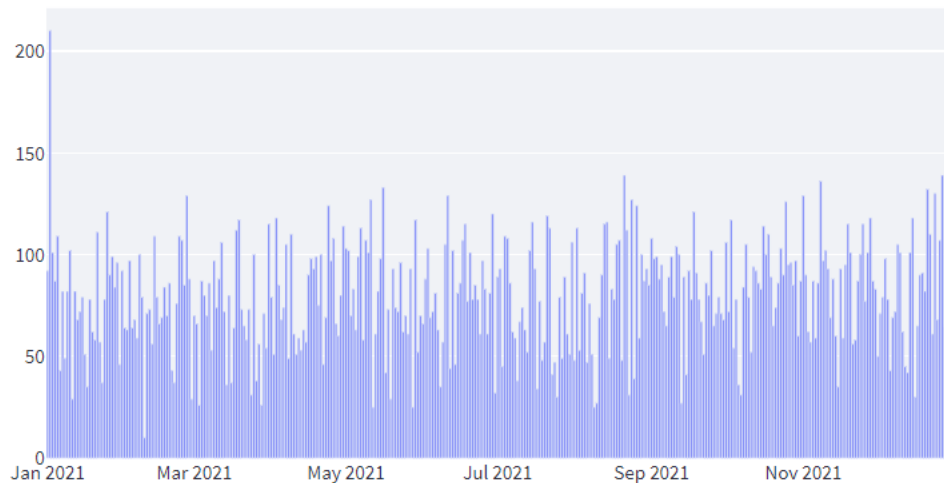


Figura 16 Shitjet e produktit gjatë muajve për vitin e caktuar

Figura 16. Në figurë është paraqitur një graf ku tregon shitjet e produktit të caktuar që vet përdoruesi e ka përzgjedhur për vitin që ata e kanë përzgjedhur. Ata e kanë mundësinë që të shohin shitjet jo vetëm për muaj por edhe për çdo ditë ku vetëm mund të e zvarritin miun nëpër graf dhe mund të shohin shitjet në numër.

Muaji me me se shumti shitje per produktin: Golden Eagle eshte 2021-01-02 00:00:00

Figura 17 Muaji për më së shumti shitje për produktin

Figura 17. Në këtë figurë jam munduar që në fund të krejt këtij procesi të ja lehtësoj pronarëve dhe menaxherëve të marketeve dhe supermarketeve që të e kenë më të çartë se në cilën datë ka pasur më së shumti shitje për atë vit që ata e kanë zgjedhur të shohin. Në këtë rast kam paraqitur datën se kur ka pasur më së shumti shitje për produktin “Golden Eagle” e që është data 02/01/2021, që i bie të jetë muaji Janar.

## 5. Konkluzioni

Në këtë studim, ne synuam të hetojmë teknikat që mund të ndikojnë në shitjet e produkteve dhe të përcaktojmë algoritmin më të besueshëm për parashikimin e shitjeve. Parashikimet tona treguan se algoritmet me potencial për të ndikuar në shitjet e produktit janë Random Forest dhe Decision Tree Regression.

Megjithatë, bazuar në rezultatet e saktësisë, Decision Tree Regression doli si algoritmi më i besueshëm për këtë punim parashikimi.

Për më tepër, analiza jonë zbuloi se Python është gjuha e preferuar për zbatimin e këtyre algoritmeve për shkak të librarive të saj të fuqishme si NumPy dhe Streamlit. Disponueshmëria e këtyre bibliotekave rrit aftësitë e përpunimit të të dhënave dhe lehtëson zhvillimin e modeleve efikase parashikuese. Prandaj, konkludojmë se Python është teknologjia e rekomanduar për kryerjen e analizave të parashikimit të shitjeve.

Duke marrë parasysh gjetjet, ne rekomandojmë përdorimin e algoritmit të Decision Tree Regression të zbatuar në Python për të parashikuar shitjet e produkteve. Ky kombinim siguron parashikime të sakta duke shfrytëzuar avantazhet e bibliotekave të gjera të Python.

Duke adoptuar këtë qasje, bizneset mund të marrin vendime të informuara në lidhje me menaxhimin e inventarit dhe optimizimin e përgjithshëm të shitjeve.

Është e rëndësishme të theksohet se kërkimet dhe eksperimentet e mëtejshme inkurajohen për të eksploruar algoritme dhe teknika shtesë që mund të ndikojnë gjithashtu në shitjet e produktit. Për më tepër, përparimet në teknologji dhe disponueshmëria e bibliotekave të reja mund të kërkojnë rishikimin e zgjedhjes së algoritmeve dhe gjuhëve të programimit në të ardhmen.

## 6. Rekomandime

Për të përmirësuar rezultatin e shitjeve të biznesit tuaj sipas analizës sonë duke përdorur algoritme të mësimit të makinerive brenda raportit tonë të modelit të parashikimit; ne rekomandojmë:

1. Ngritja e vlerësimeve të klientëve: Studimi ynë konfirmoi se vlerësimi i konsumatorëve qëndron si parashikuesi më i madh i suksesit të shitjeve. Duke i dhënë përparësi cilësisë së produktit superior dhe ofrimit të shërbimeve të jashtëzakonshme për konsumatorin, do të rrisë nivelin e përgjithshëm të kënaqësisë duke çuar në vlerësime të favorshme nga klientët.

2. Vërtetoni strategjitë e çmimeve: Hetimi renditi përfundimisht pikën e çmimit si përcaktuesin e dytë të rritjes së numrit të shitjeve. Kompanitë duhet të studiojnë modelet e tregut, së bashku me vlerësimin e taktikave të çmimeve të konkurrentëve, duke i përafruar këta faktorë me kërkesat e konsumatorëve.

3. Bazuar në hulumtimet, ofertat promociionale janë treguar se ndikojnë mesatarisht në shitje. Prandaj, bizneset duhet të konsiderojnë ofrimin e promovimeve joshëse dhe të përshtatshme për të tërhequr klientët dhe për të rritur shitjet.

4. Bizneset që kërkojnë të optimizojnë parashikimet e tyre të shitjeve mund të përfitojnë nga përdorimi i algoritmeve të mësimit të makinerive. Analiza jonë zbuloi se algoritmi Regresor i Pemës së Vendimit shfaqti saktësi superiore në parashikimin e rezultateve të shitjeve. Prandaj mund të ia vlen të hulumtohet më tej kjo metodë për zbatimin e mundshëm.

Çelësi për arritjen e parashikimeve të sakta dhe vendimmarrjes optimale qëndron në mbajtjen e ritmit me ndryshimin e kushteve të tregut dhe shijeve të klientëve.

Si të tilla ndërmarrjet duhet të mbajnë një qasje dinamike drejt modelit të tyre të parashikimit të shitjeve, duke e përditësuar vazhdimisht atë për të qëndruar të rëndësishme në peizazhin e sotëm konkurrues. Duke adoptuar këtë qasje, bizneset mund të zhbllokojnë performancën e përmirësuar të shitjeve së bashku me përfitim më të lartë.

## Referencat

- Brown, A. (2020). Demand Prediction.
- C, N. (2021, 04 14). *activestate*. Gjetur në top-10-python-machine-learning-packages: <https://www.activestate.com/blog/top-10-python-machine-learning-packages/>
- Chen, S. S., & Chen, X. . (2021). Ridge regression with quadratic loss for high-dimensional data. *Journal of Multivariate Analysis*.
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2021). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 11-37.
- Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. *O'Reilly Media, Inc.* O'Reilly Media, Inc.
- Goodfellow, I., Bengio, Y., & Courville, A. (2021). Deep learning. *MIT Press*.
- Gupta, P. &. (2021). Comparative Analysis of Machine Learning Algorithms for Sales Forecasting. *11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, (fv. 52-56).
- Hoerl, A. E., & Kennard, R. W. . (2020). Ridge regression. *Wiley StatsRef: Statistics Reference Online*.
- Hunter, J. D. (2020). Matplotlib: A 2D graphics environment. *IEEE Annals of the History of Computing*, 21-29.
- Johnson, R. (2018). Sales Forecasting.
- Lee, S. (2021). Price Optimization.
- Liu, H., Motoda, H., & Yu, H. (2020). Feature selection for knowledge discovery and data mining. *Springer*.
- McKinney, W. (2020). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51-56.
- Mohan, A., & Kumar, N. (2021). Impact of customer reviews and ratings on sales: An empirical investigation. *Journal of Retailing and Consumer Services*,, 59.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2020). Introduction to linear regression analysis. *John Wiley & Sons*.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.



- Nederkoorn, C. (2021, 04 14). *top-10-python-machine-learning-packages*. Gjetur në activestate: <https://www.activestate.com/blog/top-10-python-machine-learning-packages/>
- Oliphant, T. E. (2020). A guide to NumPy. *USA: Trelgol Publishing*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2021). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825-2830.
- Rahimi, A., & Recht, B. . (2021). Random features for large-scale kernel machines. *Journal of Machine Learning Research*, 1-44.
- Rajput, A., & Kaur, M. . (2020). Applications of Python in Machine Learning. In *Data Analytics and Learning*. Springer, 203-214.
- Rajput, R., & Kaur, R. . (2021). Machine Learning using Python: A Comprehensive Guide. *Springer International Publishing*.
- Sharda, R., Delen, D., Turban. E. (2021). *ANALYTICS, DATA SCIENCE, & ARTIFICIAL INTELLIGENCE*. Pearson.
- Teguh, A., Setiawan, N. A., & Santoso, I. (2021). Classification of fruits using artificial neural network with python. *nternational Journal of Computer Science and Information Security (IJCSIS)*, 79-86.
- Waskom, M., Botvinnik, O., O’Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., ... Seabold, S. . (2021). mwaskom/seaborn: v0.11.2. *Zenodo*.
- Waskom, M., Botvinnik, O., O’Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., ... & Sievert, S. (2021). mwaskom/seaborn: v0.11.1. *Zenodo*. .
- Williams, L. (2019). Market Analysis.
- Yoon, H., & Song, B. (2021). A Machine Learning Approach for Sales Prediction: A Comparative Study between Random Forest and Decision Tree. *Symmetry*.

## Shtojca A

```
import streamlit as st
import pandas as pd
import numpy as np
from sklearn.tree import DecisionTreeRegressor
import plotly.graph_objs as go

# Leximi i datasetit pra leximi i te dhenave qe ne kemi
df = pd.read_csv("dataset_streamlit.csv", parse_dates=['Data'])

#Vendosja e titullit per aplikacionin tone
st.title("Parashikimi i shitjeve te produkteve")

# Sidebar- Zgjedhja e produktit qe ne duam te shohim
st.sidebar.header('Zgjidhni produktin')
product_names = df['Produkti'].unique()
selected_product = st.sidebar.selectbox('Zgjidhni një produkt',
product_names)

# Sidebar - Zgjedhja e vitit qe ne duam nga 2021-2023
st.sidebar.header('Zgjidhni vitin')
years = ['2021', '2022', '2023']
selected_year = st.sidebar.selectbox('Zgjidhni një vit', years)

# Filtro të dhënat sipas produktit dhe vitit
selected_product_data = df[(df['Produkti'] == selected_product) &
(df['Data'].dt.year == int(selected_year))]

# Konvertoni KohaDites në kohë
selected_product_data['KohaDites'] =
selected_product_data['KohaDites'].map({'Mengjes': '08:00-12:00',
'Mesdite': '13:00-17:00', 'Mbremje': '18:00-22:00'})

# Grupimi i te dhenave per KohaDites
grouped_data_koha = selected_product_data.groupby(['KohaDites'],
as_index=False)['Shitjet'].sum()

# Shfaqja e plot/char per KohaDites
koha_fig = go.Figure()
```

```

koha_fig.add_trace(go.Bar(x=grouped_data_koha['KohaDites'],
y=grouped_data_koha['Shitjet']))

koha_fig.update_layout(title=f'Parashikimi i shitjeve per kohen e
dites per produktin: {selected_product} ne vitin: {selected_year}')

st.plotly_chart(koha_fig)

# Grupimi i te dhenave per muaj

grouped_data_month = selected_product_data.groupby(['Data'],
as_index=False)['Shitjet'].sum()

# Me e gjet ne cilin muaj produkti i caktuar ka pasur shitjen me te
mire

best_month =
grouped_data_month.loc[grouped_data_month['Shitjet'].idxmax(), 'Data']

# Shfaqja e plot/char per muajin

month_fig = go.Figure()

month_fig.add_trace(go.Bar(x=grouped_data_month['Data'],
y=grouped_data_month['Shitjet']))

month_fig.update_layout(title=f'Parashikimi i shitjeve per muajin e
vitet per produktin: {selected_product} ne vitin:
{selected_year}')

st.plotly_chart(month_fig)

st.write(f"Muaji me me se shumti shitje per produktin:
{selected_product} eshte {best_month}")

# Perdorimi i modelit Decision Tree

X = pd.DataFrame(selected_product_data['Data'])
y = pd.DataFrame(selected_product_data['Shitjet'])

regressor = DecisionTreeRegressor(random_state=0)

regressor.fit(X, y)

```

## Shtojca B

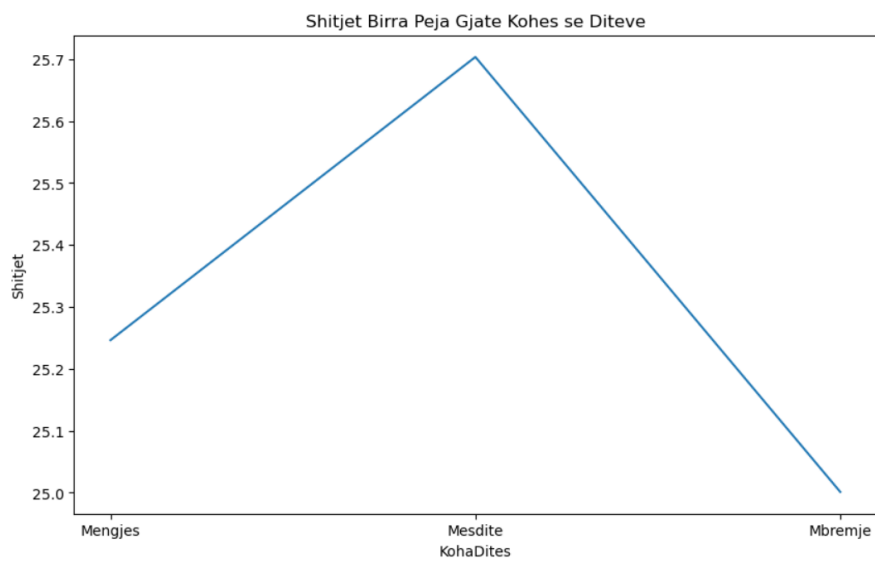


Figura 18 Shitjet Birra Peja gjatë ditës

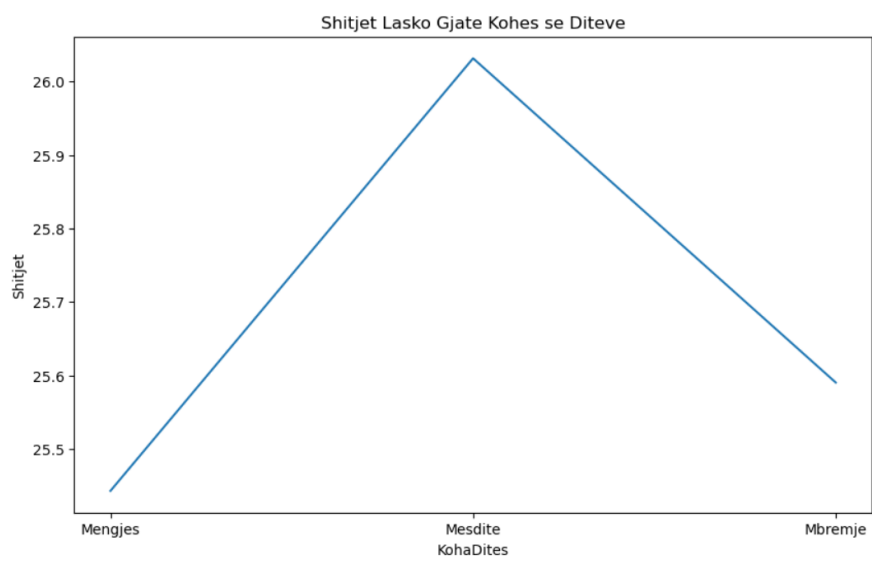


Figura 19 Shitjet Lasko gjatë ditës

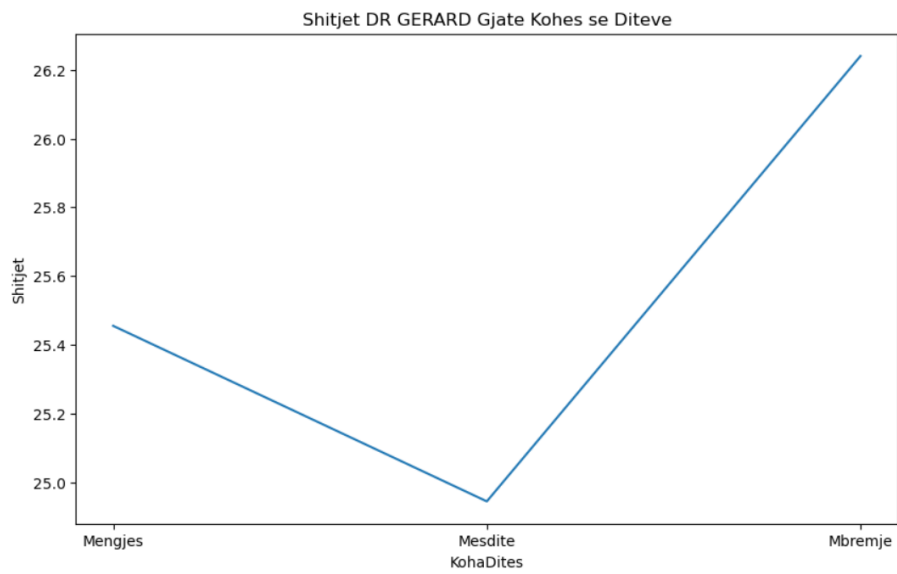


Figura 20 Shitjet DR GERARD gjatë ditës

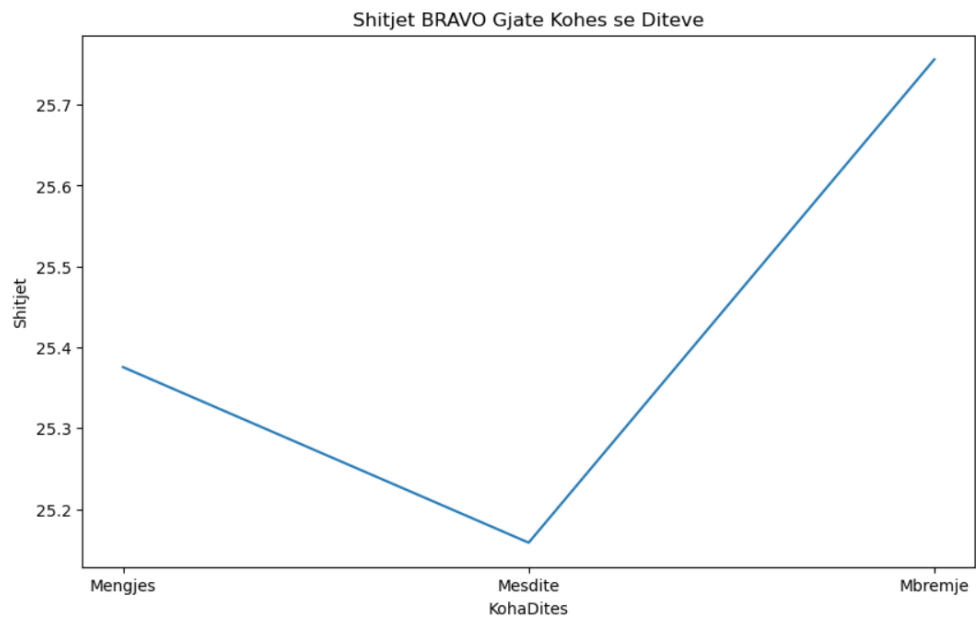


Figura 21 Shitjet BRAVO gjatë ditës