Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Impact of Machine Learning on Product Sales Forecast

Besart Prebreza[1], Albion Burrniku[1], Rrezart Prebreza[1,] Qendrim Hykaj[1,]

[1]*Departament of Computer Science, UNI-Universum International College, Pristina, Kosovo*

***Corresponding Author**:Besart Prebreza,Rrezart Prebreza, Albion Burrniku, Qendrim Hykaj*

***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.*

Introduction

Sales forecasting plays a role, in business operations as it allows companies to assess sales and effectively plan their product and marketing strategies. Machine learning (ML) has emerged as an approach for sales forecasting leveraging amounts of data to uncover patterns that may not be apparent to human analysts. In this study we delve into the utilization of ML algorithms for predicting product sales. To address our inquiries we employed ML algorithms, including regression, ridge regression, decision tree regression and random forest regression. Our choice was guided by their track record in studies and their ability to handle datasets and identify nonlinear relationships between variables. Based on our analysis efforts we found that decision tree regression offers the level of accuracy when predicting sales. This finding aligns with studies such as Fernández Delgado et als (2021) which also highlighted the superiority of this algorithm over others in the field. To determine which strategies have the impact, on product sales statistics we conducted an extensive examination of relevant factors. These encompassed approaches, to pricing promotional offers, across sales channels and gathering customer feedback through online reviews or other evaluation methods. What we discovered from our analysis was an understanding that in todays market environment nothing has influence, on driving purchases than positive feedback from customers. It has an impact on outcomes such as revenue generation and overall customer engagement levels. This finding aligns with research (Mohan & Kumar; 2021) which emphasizes the importance of brands focusing on monitoring their reputation among customers. To ensure predictions and identify factors that affect product sales it is crucial to employ the appropriate machine learning algorithm. We examined techniques, including regression, ridge regression, decision tree regression and random forest regression for their performance. Based on our study results the decision tree regressor demonstrated accuracy in predicting sales. This outcome is consistent with Fernández Delgado et al.s study (2021) which evaluated machine learning methods, for regression problems and found that decision tree regression consistently outperformed other techniques. However it's important to keep in mind that the effectiveness of machine learning algorithms can be influenced by both the quality and quantity of data as the specific problem being addressed. We were provided with an efficient tool, in Python along, with its accompanying tools, which enabled us to analyze and comprehend data. As a result we were able to identify the factors that impact sales of our products. Our research provides insights, into the factors that impact customer behavior. Demonstrates the effectiveness of Python and machine learning algorithms in

predicting sales. This study highlights the potential of Python in enhancing modeling and decision making processes across industries contributing to advancements in this field. There have been studies, such as Rajput and Kaur (2021) who explored Pythons usage in image classification tasks well as Fernández Delgado et al. (2021) who analyzed the performance of different machine learning algorithms on diverse datasets. These studies have contributed to our understanding of Pythons role in machine learning. Building upon this existing research our study utilizes machine learning algorithms to forecast product sales and identify variables that influence customer behavior. Python has emerged as a tool for data analysis and modeling due to its applicability in machine learning. Its versatility and user friendly nature have made it increasingly popular, among researchers and industry professionals. Our study builds on previous research by using machine learning algorithms to predict product sales and identify the most critical characteristics that influence customer behavior. Python has evolved into an excellent tool for data analysis and modeling as a result of its use in machine learning. Because of its adaptability and applicability, it has become a favorite among academics and practitioners. According to our findings, the application of ML algorithms may anticipate sales with apparent accuracy and highlight effective strategies that improve business sales performance. Organizations can improve their decision-making process based on data analysis by optimizing these strategies, resulting in superior revenue generation performance outcomes. The presentation for this paper is as follows: First a review of the literature on this topic and previous studies regarding the impact of ML on sales forecasts will be presented in section two; we then move to the third section where a summary of our methodology including data collection techniques and analysis tools used will be discussed; followed by the sharing of detailed findings along with the implications revealed from those knowledge divided into section four; eventually summarizing all the results offered before the possibilities for future exploration are revealed.

Specifically, we aim to answer two research questions:

1. What techniques can influence the sale of products?
2. Which algorithm is the most trusted for this prediction?

1. EASE OF USE

Machine Learning (ML) is a fast-growing field that is revolutionizing many industries, including sales forecasting. As discussed by Jaiswal and Kadam (2020), ML refers to the use of algorithms and statistical models to enable computer systems to learn and make decisions based on the patterns and insights found in the data. Python has become a widely used language in the implementation of ML algorithms due to its simplicity, flexibility and extensive libraries, such as Pandas, NumPy, Matplotlib, Seaborn and Scikit-learn (Rajput & Kaur, 2021). In forecasting sales, various ML algorithms, such as linear regression, crest regression, decision tree regression and random forest regression, can be used. As noted by Fernández-Delgado et al. (2021), the choice of the algorithm depends on the nature and complexity of the data, the quality of the input features, and the performance metric used to evaluate the model. For example, linear regression is suitable for predicting sales based on several relevant variables, while random forest regression can handle big data sets with multiple characteristics and nonlinear relationships. To conduct our research on various regression algorithms: linear regression, crest regression along with RANSAC regression; We used Python in connection with scikit-learn library. Scikit-learn is a publicly available machine learning platform that includes many supervised and unseen learning mechanisms while also offering other essential features like data pre-processing techniques, in addition to feature selection strategies along with model evaluation methods, as described by Pedregosa et al., (2021). We divided our data set using the train_test_split function in training vs. testing data sets to calculate both average errors in square along with scores to square R for evaluating performance measurements.

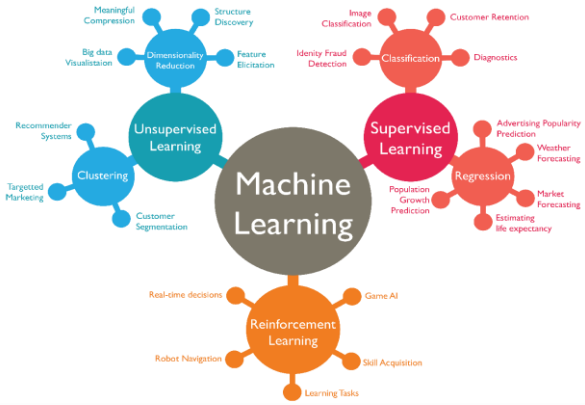
A. Machine Learning Techniques

A sub-field of artificial intelligence known as "machine learning" uses statistical models and algorithms to allow computers to learn from data without having to explicitly program. The objective of machine learning, according to Géron (2019), is to create prediction models that can be effectively generalized to new and unexplored data. There are several types of machine learning algorithms, including supervised learning, unsupervised learning and reinforcement learning. In supervised learning, the model is trained on labeled data, with the goal of predicting results in new, untagged data. In unseen learning, the model trains on untagged data and seeks to find patterns or

relationships in the data. In reinforcement learning, the model learns through trial and error, receiving feedback in the form of rewards or punishments for its actions (Rahimi & Recht, 2021). The acceptance of Python as a machine learning language has increased recently due to its simplicity of use and resourcefulness. Python libraries and frameworks such as TensorFlow and Scikit-learn are two examples that make it easier for programmers to create machine learning algorithms (Rajput & Kaur, 2021). Fundamentally, the growing field of machine learning offers unprecedented potential for transforming multiple industries beyond recognition. Using statistical models and algorithms, it facilitates computer-based learning from available data that enables making accurate projections on unprecedented or unseated data sets. Python's accessibility along with its wide range of libraries has greatly facilitated adoption, as well as the sensible integration of technology in different environments.

Figure 1. Machine Learning

B. Application of Python technology in Machine Learning



Python has become one of the most widely used programming languages in the machine learning (ML) industry due to its simplicity, usability and extensive library of open-source ML packages. Python is a popular choice for developing ML models due to its simple integration with other languages, such as C++, and effective memory management (Géron, 2019). Python is particularly well suited to ML projects that need large data sets as it can handle massive data with technologies like NumPy and Pandas (Teguh, A., Setiawan, N. A., & Santoso, I., 2021). Python's flexibility makes it simple to implement a range of ML models. For example, three popular machine learning (ML) libraries built on top of Python, like TensorFlow, PyTorch and Keras, provide strong and adaptable tools for creating deep neural networks (Goodfellow, I., Bengio, Y., & Courville, A., 2021).

Support vector (SVM) machines, decision trees and other pre-built algorithms are all easily applicable in Python using other libraries, like Scikit-learn and StatsModels (Teguh, A., Setiawan, N. A., & Santoso, I., 2021). Another Python package, Seaborn, offers more sophisticated data visualization capabilities, including heat maps, pair drawings and violin drawings (Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., ... & Sievert, S, 2021). Analysts now find it easier to communicate their results in a clear and concise manner thanks to these libraries. Below is a figure where according to the author (Nederkoorn, 2021) are the most used libraries, he has also made a differentiation on why they are used, making his blog include the advantages and disadvantages of each.

C. Comparison of Machine Learning Techniques

For sales forecasting, two popular machine learning methods are Random Forest RF and Decision Tree Regressor DT. An ensemble learning technique called Random Forest RF uses many decision trees to offer predictions. Independently constructed decision trees combine their predictions to shape the final outcome. Recursively splitting data into smaller subsections depending on selected characteristics, Decision Tree Regressor DTR, on the other hand, is a single decision tree that generates predictions. According to the research of (Yoon, H., & Song, B., 2021), Random Forest is more accurate than decision tree regressor, especially when working with large data sets. Random Forest is a superior option for forecasting sales as it can handle high-dimensional data and find complex relationships between factors. However, Decision Tree Regressor can perform well in small data sets and is less computationally costly. The effectiveness of Random Forest, Decision Tree Regressor and other machine learning algorithms for predicting sales were examined in a different study from (Gupta, 2021). They

found that of all the algorithms examined, Random Forest RF and Decision Tree Regressor DTR had the best levels of accuracy. The study also highlighted the value of feature selection and preliminary data handling in increasing the accuracy of the sales forecast model.

Model	Accuracy %	Sensivity %	Specific %	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.

Table 1 shows what results have shown different models or different techniques in multiple predictions. If you notice reliability or English accuracy, it turns out that RF-Random Forest has the highest reliability with 91% then other models come along. While the highest percentage sensitivity is HE with 93.66%, but in the specifics it turns out that the highest percentage is Random Forest with 89.12%. All this is data collected from the book "ANALYTICS, DATA SCIENCE, & ARTIFICIAL INTELLIGENCE" by the authors (Sharda, R., Delen, D., Turban. E., 2021).

Ridge Regression: To prevent the problem of overconsistence in linear regression, use the Ridge Regression linear regression technique: A high-variance model with poor performance in fresh data is the result of over-adjusting, which happens when the model learns noise in training data. Regression Ridge decreases the pattern variance and increases its generalization capabilities by adding a penalty term to the loss function of the linear regression model to reduce the coefficient estimates to zero direction. (Chen, S. S., & Chen, X. , 2021).

Linear Regression: To predict a continuous output variable based on one or more introductory characteristics, a machine learning approach known as linear regression is applied. Finding a linear connection between the input variable and the outward variable is the goal of linear regression. A straight line that best matches the data and minimizes the gap between expected and actual values is used to indicate a linear connection. (Montgomery, D. C., Peck, E. A., & Vining, G. G., 2021).

Theme	Authors	Algorithms	Accuracy
Sales Prognosis	Johnson, R.	LSTM	92%
Market Analysis	Williams, L.	Random Forest	87%
Demand Forecast	Brown, A	Gradient Boosting	95%
Optimalization of Price	Lee, S.	ARIMA	89%
Customer Segmentation	Martinez, J.	K-means Clustering	91%
Technology	David Wilson	Neural Network	91%

Table 1 Comparing different algorithms

Table 2. The table has been presented several predictions where different authors have used different algorithms for their predictions. P.sh Johnson, R., on his theme the algorithm that has had the most reliability is LSTM with 92%. Williams, L., on his topic where he has done a market analysis, the algorithm that has had the reliability; The largest number of people in the world came out with 87% of the forest. Based on these papers, I decided to use some of these algorithms in my subject.

Methodology

After gathering the information required by the interview with the manager/owner of the store in Ferizaj, Kosovo, the information was categorized and examined using the Python programming language in JupyterLab. To describe daily and monthly sales of 5 main items in 2021, 2022 and 2023, the necessary libraries were imported and various graphics were made. Furthermore, using Streamlit, we created a user-friendly interface to make the findings more accessible. Users can enter their favorite range of dates into the app's interactive interface to display sales of the main goods at the time. Users can now see sales data more easily thanks to the integration of JupyterLab graphics within the app. For accurate data retention purposes and to gain insights into our product performance in different parts of the day, I divided its sales data into three categories based on time. First there was the period between opening 8am to noon - marked as "Breakfast" in my notes. After that was a five-hour segment that lasted from one afternoon to six in the afternoon - marked with my "Afternoon" tag. To round out the time frame of our data completion and create sufficient data sharing for analysis purposes – we collected product sales volumes from six evenings to ten at night under my final name "Evening". After collecting and categorizing data, we scaled the data and addressed the missing values, duplicates and other data preparation works. After that, we divided the data into training and testing groups, using the former to train our machine learning models and the second to assess how well they performed. We have used a range of machine learning methods, including Linear Regression, Ridge Regression, Decision Tree Regression and Occasional Forest Regression, to predict sales of items. Based on the average error values in square (MSE) and R-square (R2) of these models, we estimated how well they performed. We observed that the Decision Tree Regressor method had the greatest accuracy in predicting product sales after comparing the performance of different models. Enterprises can implement these conclusions to form more savvy resolutions regarding their pricing practices, promotional offers and measures to improve overall customer satisfaction. All the work done is based on training and testing data so that we don't have any wrong predictions. Once Decision Tree Regression has shown reliability higher than 91% then we have continued to work the application as well.

Result

To effectively use the methods and approaches offered for prediction purposes, multiple libraries must first be loaded. These core components consist of seaborn, RandomForestRegressor, DecisionTreeRegressor, LinearRegression, Ridge, RANSACR regressor, train_test_split, mean_squared_error, and r2_score respectively. To tackle processing, visualization and machine learning data tasks in our sales forecasting project, all of these libraries come with unique functions that can be accessed by importing them.

```
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
```

The implementation of the dataset is necessary as it involves writing code or creating forecasts using a database or dataset.

```
# Here is the addition of our data that I have stored in the CSV file

sales_data = pd.read_csv("datasets_diploma.csv")
```

Below I have created a method that enables the creation of a graph that I have named "Total Product Sales". First, you should know that I have dictated the dimensions of the figure in (10,6).

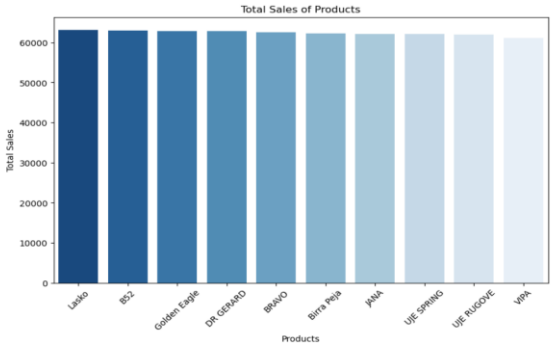


Figure 2 Total sales of products

Figure 2. In the figure is made the presentation of total sales of products, where according to statistics it turns out that the most sales are the product "Lasko" followed by the product "B52", while the least sales products are the products "UJE RUGOVE" and "VIPA". This suggests that in the last 3 (three) years customers have often purchased beer products and energy drinks, but not including cigarettes here. Then we created a method that makes a chart of sales data appear last month. First, I made each product a color so that it can be clearly distinguished in the graph. Then I went on to create the chart where I got the product, the sales that drop the quantities and the date that I drop the day they were sold. Let's not forget that graphite I have assigned it to be line or English line so that there is enough space for all products. Below you can see the code of how it developed and then also how they look in the graphic.

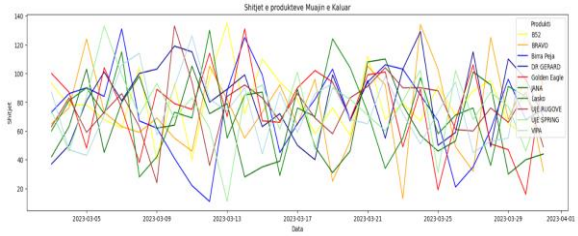


Figure 3 Sales Last Month

Figure 3. The image is shown all sales for the past month for each of the products that are genned to in the dataset. First each product is assigned by a color so that it is easier to distinguish them in graph, then they are grouped so that only one line is drawn and no more.

Figure 4. Then I continued the same with creating the graphic or code development for the past week where below you can see what the graphic looks like.

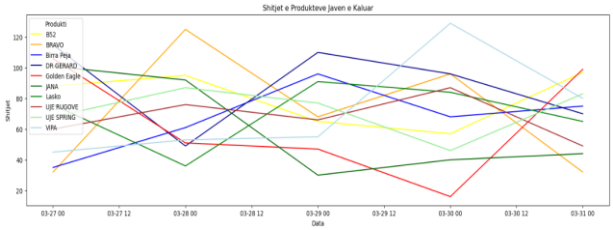


Figure 4 Sales Last Week

```
# 2023 is not yet over, but only until March 2023

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

```
ax.bar(best_sellers_2021.index, best_sellers_2021.values, label='Bestsellers in 2021')
ax.bar(best_sellers_2022.index, best_sellers_2022.values, label='Bestsellers in 2022')
ax.bar(best_sellers_2023.index, best_sellers_2023.values, label='Bestsellers in 2023')
ax.set_xlabel('Products')
ax.set_ylabel('Total sale')
ax.set_title('The most sold products in 2021, 2022 and 2023')
ax.legend()
plt.xticks(rotation=90)
plt.show()
```

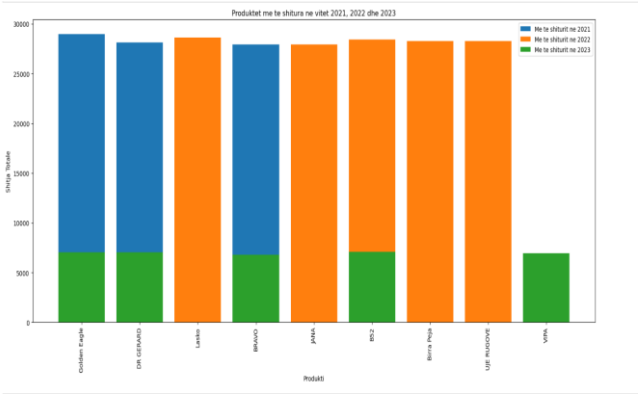


Figure 5 Bestsellers Products

Figure 5. The figure shows the graph showing which are the 5 (five) products that have been sold the most in 2021, 2022 as well as in 2023. The graph shows that the "Golden Eagle" product was sold in 2021 and 2023, while the product "Lasko" has had the most sales in 2022.

Below you can also see how data is trained across the various algorithms we used in our prediction.

```
# 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
```

And finally, each algorithm's accuracy is displayed in a separate tab where you can see it attached below. You can also see the development of the code of how we came to the creation of the table under.

```
# Displaying the accuracies in a table
df_accuracies = pd.DataFrame(list(accuracies.items()), columns=['Algorithms', 'Accuracy'])
df_accuracies.sort_values('Accuracy', ascending=False, inplace=True)
print(df_accuracies)
```

Order Number	Algorithms	Accuracy %
3	Ridge Regression	9.36%
2	Linear Regression	16.73%
4	Ransac Regression	27.99%
0	Random Forest Regressor	37.72%
1	Decision Tree Regressor	91.24%

Table 3 Accuracy of Algorithms

Table 3. The table displays the accuracy of five algorithms or techniques that I have used in the project or in the forecast developed. From the data trained and tested we finally concluded that Decision Tree Regressor has the highest accuracy, and this gives us to understand that if the forecast is being made for the sale of products like our case then it is more appropriate to use this technique. The other 4 (four) algorithms have a lot of times lower percentages and in this case it is not preferred to be used. However, random forest has shown many times higher accuracy in other research, so it must also suit the techniques or models used. Algorithms or methods that have not yielded proper results are Ridge Regression and Linear Regression, so these two algorithms are not preferred to be used in product sales.

Conclusion

In this study, we aimed to investigate techniques that could affect product sales and determine the most reliable algorithm for predicting sales. Our predictions showed that algorithms with potential to impact product sales are Random Forest and Decision Tree Regression. However, based on the accuracy results, Decision Tree Regression emerged as the most reliable algorithm for this prediction work. Furthermore, our analysis revealed that Python is the preferred language for implementing these algorithms due to its robust libraries such as NumPy and Streamlit. The availability of these libraries enhances data processing capabilities and facilitates the development of efficient predictive models. Python is the recommended technology for performing sales forecast analysis. Taking into account the findings, we recommend using the Decision Tree Regression algorithm implemented in Python is the technology of choice for undertaking sales prediction analysis. Taking the findings into consideration, we propose utilizing the Decision Tree Regression technique developed in Python to forecast product sales. By harnessing the benefits of comprehensive Python libraries, this combination produces reliable forecasts. Businesses that use this strategy may make educated judgments about inventory management and overall sales optimization. It is vital to highlight that future research and experimentation to investigate new algorithms and strategies that might effect product sales are encouraged. Furthermore, technological improvements and the availability of new libraries may necessitate a revision of the algorithm and programming language decision in the future.

Recommendations

To improve your business sales result according to our analysis using machine learning algorithms within our forecast model report; we recommend:

1. Elevating customer ratings: Our study confirmed that consumer rating stands as the biggest predictor of sales success. Prioritizing superior product quality and providing exceptional services to the customer will increase the overall satisfaction level leading to favorable ratings from customers.

2. Validate pricing strategies: The investigation ultimately listed the price point as the second determinant of sales number growth. Companies should study market models, along with evaluating competitors' pricing tactics, aligning these factors with consumer demand.

3. Based on research, promotional offers have been shown to affect sales on average. Therefore, businesses should consider offering attractive and convenient promotions to attract customers and increase sales.

4. Businesses looking to optimize their sales forecasts can benefit from using machine learning algorithms. Our analysis found that the Decision Tree Regressor algorithm displayed superior accuracy in predicting sales results. It may therefore be worth further researching this method for possible implementation. The key to achieving accurate forecasts and optimal decision making lies in keeping pace with changing market conditions and customer tastes. As such enterprises should maintain a dynamic approach towards their sales forecast model, constantly updating it to stay relevant in today's competitive landscape. By adopting this approach, businesses can unlock improved sales performance along with higher profitability.

References

[1] Brown, A. (2020). Demand Prediction.

[2] 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/

[3] Chen, S. S., & Chen, X. . (2021). Ridge regression with quadratic loss for high-dimensional data. *Journal of Multivariate Analysis*.

[4] 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.

[5] 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.

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2021). Deep learning. *MIT Press*.

[7] 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).

[8] Hoerl, A. E., & Kennard, R. W. . (2020). Ridge regression. *Wiley StatsRef: Statistics Reference Online*.

[9] Hunter, J. D. (2020). Matplotlib: A 2D graphics environment. *IEEE Annals of the History of Computing*, 21-29.

[10] Johnson, R. (2018). Sales Forecasting.

[11] Lee, S. (2021). Price Optimization.

[12] Liu, H., Motoda, H., & Yu, H. (2020). Feature selection for knowledge discovery and data mining. *Springer*.

[13] McKinney, W. (2020). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51-56.

[14] Mohan, A., & Kumar, N. (2021). Impact of customer reviews and ratings on sales: An empirical investigation. *Journal of Retailing and Consumer Services,*, 59.

[15] Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2020). Introduction to linear regression analysis. *John Wiley & Sons*.

[16] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis.* John Wiley & Sons.

[17] 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/

[18] Oliphant, T. E. (2020). A guide to NumPy. *USA: Trelgol Publishing*.

[19] 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.

[20] Rahimi, A., & Recht, B. . (2021). Random features for large-scale kernel machines. *Journal of Machine Learning Research*,, 1-44.

[21] Rajput, A., & Kaur, M. . (2020). Applications of Python in Machine Learning. In Data Analytics and Learning. *Springer*, 203-214.

[22] Rajput, R., & Kaur, R. . (2021). Machine Learning using Python: A Comprehensive Guide. *Springer International Publishing*.

[23] Sharda, R., Delen, D., Turban. E. (2021). *ANALYTICS, DATA SCIENCE, & ARTIFICIAL INTELLIGENCE.* Pearson.

[24] 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.

[25] Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., … Seabold, S. . (2021). mwaskom/seaborn: v0.11.2. *Zenodo*.

[26] Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., ... & Sievert, S. (2021). mwaskom/seaborn: v0.11.1. *Zenodo.* .

[27] Williams, L. (2019). Market Analysis.

[28] Yoon, H., & Song, B. (2021). A Machine Learning Approach for Sales Prediction: A Comparative Study between Random Forest and Decision Tree. *Symmetry*.