

Avocado Price Prediction

Ayush Singh¹, Aadarsh Khatiwada², and Navdeep Sharma³

¹Department of Data Science and Analytics, Central University of Rajasthan

January 22, 2023

Abstract

Avocado price prediction is a valuable topic of research due to its relevance in the modern world. This research report aims to investigate the price prediction of avocados using linear regression, XGBoost, and Random Forest. The problem statement is first clarified, followed by a detailed literature review of the subject. The findings are then presented, which will be followed by a discussion of the identified shortcomings and potential avenues for further research. Finally, the conclusion of the analysis will be presented. All of the topics discussed in this report are presented with the goal of providing a thorough and accurate assessment of avocado price prediction.

1 Introduction

One of the most popular fruits in the world, avocado's popularity has grown throughout time. As a result, the price of avocados has increased in volatility and grown more unpredictable. Finding the most accurate model and conducting data analysis are crucial steps in effectively forecasting avocado prices. In this project, the price of avocado is predicted using the algorithms of linear regression, XGBoost, and Random Forest.

2 Methodology

Data collection and model training are the two processes that make up the technique for machine learning algorithms used to predict avocado prices.

Data must first be acquired from a variety of sources, including market pricing, weather predictions, consumer buying trends, and more. To

make sure the data is acceptable for the model training, it must be cleaned and preprocessed after it has been collected.

The data is then used to train an appropriate machine learning system. For this, a variety of methods can be applied, including linear regression, XGBoost, and random forests. It is crucial to pick the algorithm that best fits the available data and the intended result. After training, the model can be used to forecast the price of avocados in the future.

3 Problem statement

Statement of the Issue for Predicting Avocado Prices Using Various Machine Learning Algorithms Due to the recent increase in avocado prices in the US, forecasting algorithms that could project future costs are now necessary. In this research, multiple machine learning algorithms are developed and compared in order to create a model that can precisely forecast avocado price changes. The prices of avocados in the US from 2015 to 2018 will be used as the dataset. Total volume, total bags, small bags, large bags, and kind will all be variables in the dataset. The study will test a number of algorithms, including linear regression. Root mean square error will be used to compare each algorithm's performance (RMSE). The most effective algorithm will be used to build a model to accurately predict avocado prices

4 About the data

Context Millennials adore avocado toast, as is generally known. The fact that every millennial lives in their parents' basement is also widely recognised. They are buying way too much avocado toast, therefore it is obvious they aren't buying a house! However, there may be hope if a millennial could choose a city with inexpen-

sive avocados and pursue their American Dream. Content This data was downloaded in May 2018 as a single CSV file from the Hass Avocado Board website. The Hass Avocado Board describes the information on their website as follows: Weekly 2018 retail scan data for National retail volume (units) and price are shown in the table below. Based on real retail sales of Hass products, retail scan data is obtained directly from retailers' cash registers. The table below displays an enhanced, multi-outlet retail data set as of 2013. Grocery, mass, club, drug, dollar, and military channels are all combined in multi-outlet reporting. Even when several units (avocados) are sold in bags, the Average Price (of avocados) in the table shows a cost per unit (per avocado). Only Hass avocados are included in the table's Product Lookup Codes (PLUs). The table does not include other types of avocados, such as greenskins.

Several relevant dataset columns include:

Date - The day the observation was made. AveragePrice is the average cost of either conventional or organic avocados. calendar year Region: the city or area where the observation was made. Total Volume - Total Avocado Sales 4046 - Total Avocado Sales of 4046 with PLU The total number of avocados with PLU that were sold was 4225. The total number of avocados with PLU that were sold was 4770.

5 Real life use case

The prediction model would be trained on a dataset of historical avocado prices and other relevant factors, such as location and seasonality. The model could be used to determine the relationships between the different features, such as the price of avocados in the market, weather conditions, and the time of year. It would then make predictions about the price of avocados based on these features. It could be used to make more accurate predictions about the price of avocados, as it would be able to learn more complex relationships between the different features.

6 Results

The avocado price prediction model evaluated using linear regression, XGBoost, PCA .XGBoost performed to be best. Accuracy was measured and compared.

LinearRegression	60.18%	
RandomForestRegressor	88.13%	article
XGBRegressor	88.95%	

7 Discussion

7.1. Machine Learning Project for Avocado Price Prediction

Since avocado prices have increased recently, many consumers are interested in forecasting these prices to aid in smarter purchasing choices. In order to achieve this, a Machine Learning (ML) project can be a useful tool for precisely forecasting avocado costs and ensuring that you obtain the greatest bargain.

7.2. Data Gathering

Data gathering is the initial step of a machine learning project. You must collect information on both the supply and demand sides of the market in order to effectively anticipate avocado pricing. This can contain information on avocado production, consumer demand, and price trends. Additionally, you can gather information on outside variables that may affect the price of avocados. such as the climate, key economic indices, and even the state of politics.

7.3. Feature engineering

The process of feature engineering, which comes after data collection, is taking the data that has been gathered and turning it into features that can be used to train the ML model. This entails developing features like standardising the data and converting categorical data into numerical values. You can also wish to add features that are connected to the outside factors, such as one for each economic indicator or type of weather.

7.4. Model Training

After the data has been prepared, the ML model can be trained. To do this, a model architecture must be developed, and the model must be trained using the labelled data. The model can then be assessed to see if it is accurate enough to produce trustworthy predictions.

7.5. Prognosis

Finally, you can forecast avocado prices using the learned ML model. After that, you may use the forecasts to guide your purchases and acquire the greatest prices on avocados. In conclusion, using ML to anticipate avocado costs can be a useful tool for improving your shopping choices. You can develop a precise and trustworthy ML model for forecasting avocado prices by gathering the required data, carrying out feature engineering, and training a model.

Conclusions

The outcomes of this machine learning study demonstrate how accurately the price of avocados can be predicted. This can be used to guide decisions in the avocado business and other sectors where prices are influenced by seasonality or local demand, among other things. Future predictions will be more accurate because to the use of straightforward models that can be adjusted to better fit the data. Overall, this experiment has shown how machine learning can be used to reliably forecast avocado prices.

Acknowledgements

Many thanks to the Hass Avocado Board for sharing this data!!
<http://www.hassavocadoboard.com/retail/volume-and-price-data> Inspiration

References

- [1] Forecasting price trends in the US avocado (Persea Americana Mill.) marketEA Evans, S Nalampang - Journal of Food Distribution Research, 2009 - core.ac.uk

.