

Data Science Project in Python on BigMart Sales Prediction

The goal of this data science project is to build a predictive model and find out the sales of each product at a given Big Mart store.

Project template outcomes

- Understanding the problem Statement
- Importing the Dataset and performing basic EDA
- Checking for the null values and describing the variables
- Imputation of the Null-Values using pivot tables
- Feature Engineering/ Creating New features
- Using seaborn to understand the contribution of the categorical values on target variables
- Using boxplot for identifying outliers
- Fixing categorical variables using Label and One hot encoding
- Applying Linear, Bayesian Regression models
- Applying ensemble bagging models like Random Forest and Bagging models
- Applying boosting models like Gradient Boosting Tree and XGboost
- Applying Neural Network model MLPRegressor
- Making function for On spot-checking and selecting the best for hyperparameter tuning
- Defining function for HyperParameter tuning
- Standardization and effect of Standardization
- Understanding Robust Scaler and Normalization
- Implementing Robust Scaler and Normalization
- Concluding the final model and predicting for the test data set
- Saving the model using Joblib

Description

Occasionally, we all love to go on a shopping spree, only to realize later that most of the things we purchase may not be precisely valuable. One of the primary reasons behind this could be store executives' clever placement of different products. You might be wondering if that is the case, how can one then protect themselves from falling for such smartly curated traps? Find out the answer to it as we discuss briefly one of the most straightforward machine learning project ideas for beginners: BigMart Sales Prediction.

Introduction to Big Mart Sales Prediction Project in Python

The BigMart sales prediction dataset contains 2013's annual sales records for 1559 products across ten stores in different cities. Such a vast amount of data can reveal insights about apparent customer preferences as specific attributes of each product and store have been defined in the dataset.

This data science project aims to build a predictive model and determine the sales of each product at a particular store. Using this model, BigMart will try to understand the properties of products and stores, which play a crucial role in increasing sales and developing better business strategies.

Tech Stack

Language: Python

Libraries: Pandas, NumPy, Matplotlib, sklearn

Data Science Concepts You Will Master in this BigMart Sales Prediction ML Project

We have briefly discussed the key learning takeaways from this python sales prediction project.

Exploratory Data Analysis

Understanding the dataset and the distribution of variables is critical in implementing any ML project ideas. Thus, the first data science concept to explore in this big mart sales prediction project is plotting graphs to understand the data. You will learn how to generate visualizations that help analyze outliers threshold and reveal variables containing missing values. Furthermore, the graphs will also assist in determining insights about different stores, for example, how the size of an outlet is related to the profit it generates. Visualizing data also helps decide thresholds for converting categorical variables into numerical values and deducing irrelevant feature variables.

Data Preprocessing

The Bigmart sales dataset contains different types of variables, and this project will assist you in processing different data types. Additionally, the data has missing values as some stores do not report all the data due to technical glitches. Hence, you get to learn how to treat them accordingly. Also, you will learn about two essential methods: label encoding and one-hot encoding for treating categorical variables. The dataset also contains many outliers, and it is not a good practice to train your model with those distant values. This big mart sales prediction machine learning project will thus help you in learning how to eliminate specific outliers from the dataset. Furthermore, you will learn how to perform and create the test-train dataset to apply machine learning algorithms.

Machine Learning Algorithms

As predicting the Big mart sales is a regression problem, this project will guide you on applying regression machine learning algorithms. You will learn about linear regression, the Bayesian model, ensemble bagging model like the random forest, boosting models, neural network, and MLP regressor. Additionally, the project will teach you about hyperparameter tuning for all the models that one can use to prepare the Bigmart sales prediction project report. This project is one of the excellent ML Projects for beginners to explore how machine learning and data science tools assist in solving large-scale business problems.

FAQs on Sales Prediction in Python

1) Which algorithm is best for sales prediction?

Various models are used for predicting sales. These include random forest, linear regression, ARIMA, LSTM, and XGBoost. Based on the given dataset, you can test any of these machine learning models for efficiency after hyperparameter tuning.

2) How does Python predict future sales?

Python contains a sklearn library that contains various regression models like the random forest, linear regression, etc., that one can use for sales prediction. Another way to predict sales is to use Time series models for prediction.