Project Title:

Food Demand Forecasting For Food Delivery Company Using IBM Cloud

Abstract:

As the food delivery services highly depend on the perishable goods,it could happen that more amount is reserved and stored for supplying then required,leading to wastage of the material and also economic losses.For a situation like this it would be very helpful if already and idea of the required amount is known in advance.Using Machine Learning Models we can predict the number of orders that can be placed considering appropriate information like area, city etc., and meal information like category of food, sub category of food, price of the food or discount in a particular week.By using this data, we can use any classification algorithm to forecast the quantity for 10 weeks.Also UI is created for the same using the html and also Flask is used,which is a web framework for server side scripting .

Problem Definition:

To predict the number of orders based on the appropriate information considered.

Goals and Objective:

- To perform the preprocessing to enhance the accuracy and efficiency of the model
- To do the data visualization of the data for its better understanding
- To choose a model that gives least Root mean square error
- Use IBM services to train the model and deploy it
- To integrate Flask with the scoring end point

Outcome:

- Getting the predicted number of orders
- Wastage of food can be reduced

Domain Description:

Machine Learning

It is the branch of AI where algorithms are used to learn from data to make future decisions or predictions.The different types of ML are Supervised,Unsupervised and Reinforcement Learning.In supervised learning,machine learns explicitly from the labelled data and data is there with clearly defined output.It is mainly used for prediction.An example for it can be spam filtering.Whereas there is no such labelled data present in unsupervised learning and the evaluation is mostly indirect and it is not used for the problem statements involving predictions.In reinforcement learning,it is basically self learning and rewards and penalty based.

IBM Services:

- **Watson :**Watson is AI from IBM. Created to make your business more intelligent and every worker your best worker. Watson features a range of advanced APIs, specialized tooling, and Software as a Service application. This implies that Watson is made for complex use cases and designed to integrate with platforms that experts utilize in their daily work. Ensuring seamless access to the knowledge you would like to form the right decisions.
- **Watson Machine Learning:**Using IBM Watson Machine Learning, you can build analytical models and neural networks, trained with your own data. You can also deploy models, scripts, and functions, manage your deployments, and prepare your assets to put into production to generate predictions and insights.

Major steps in the project:

- First the dataset is downloaded from kaggel.com
- Importing important libraries:

Scikit-learn: Scikit-learn has a wide range of supervised and unsupervised learning algorithms that works on a consistent interface in Python.The library can also be used for data-mining and data analysis.The main machine learning functions that the Scikit-learn library can handle are classification,regression,clustering,dimensionality reduction,model selection, and preprocessing.

Pandas:Pandas are turning up to be the most popular Python library that is used for data analysis with support for fast,flexible,and expressive data structures designed to work on both “relational” or “labeled” data.Pandas today is an inevitable library for solving practical real-world data analysis in Python.Pandas is highly stable, providing highly optimized performance.The backend code is purely written in C or Python.

NumPy: NumPy is a well known general-purpose array-processing package.An extensive collection of high complexity mathematical functions make NumPy powerful to process large multi-dimensional arrays and matrices.NumPy is very useful for handling linear algebra,Fourier transforms, and random numbers.Other libraries like TensorFlow uses NumPy at the backend for manipulating tensors.

Matplotlib: Matplotlib is a data visualization library that is used for 2D plotting to produce publication-quality image plots and figures in a variety of formats.The library helps to generate histograms,plots,error charts,scatter plots,bar charts with just a few lines of code.

Seaborn: Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis.

Pickle: Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk.

- Data Preprocessing-This step involves checking for any null or missing values,merging the csv files if required,checking for any column that is not required and then dropping that column.One of the major step is Label encoding.Label Encoding refers to converting the labels into a numeric form so as to convert them into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. It is an important preprocessing step for the structured dataset in supervised learning.
- Data Visualization is done.Machine learning data visualization is important to understand

how data is used in a particular machine learning model it helps in analyzing it.

- Data is split into train and test data
- Various machine learning models are applied and the one which gives the minimum root mean square error is chosen.The various machine learning models used are:

1. Linear Regression: Linear Regression is a linear model that assumes a linear relationship between the input variables (X) and the single output variable (y). It’s particularly useful when the relationship to be modeled is not extremely complex and if you don’t have a lot of data.

2. Lasso Regression:LASSO regression is a variation of Linear Regression that uses Shrinkage. Shrinkage is a process in which data values are shrunk towards a central point as the mean. This type of regression is well-suited for models showing heavy multicollinearity (heavy correlation of features with each other).

3. ELasticNet Regression / Classification: It’s a hybrid of Lasso and Ridge Regression techniques, therefore it’s also well-suited for models showing heavy multicollinearity (heavy correlation of features with each other).

4. Decision Tree Regression / Classification:Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

5. KNeighors Regressor:KNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighbourhood. The size of the neighbourhood needs to be set by the analyst or can be chosen using cross-validation to select the size that minimises the mean-squared error.

6. Gradient Boosting Regressor:Gradient Boosting is a popular boosting algorithm. In gradient boosting, each predictor corrects its predecessor's error. In contrast to Adaboost, the weights of the training instances are not tweaked, instead, each predictor is trained using the residual errors of the predecessor as labels.

7. XGB Regressor:XGBoost is a powerful approach for building supervised regression models. The validity of this statement can be inferred by knowing about its (XGBoost) objective function and base learners.

- The best rmse model is saved,in this case it is the decision tree.Locally it is saved using the pickle library.
- The model can be trained on the ibm watson.We can create a deployment space and save the model there.After that we can deploy that model and integrate Flask with its scoring end point
- For the front end design html is used.Home.html is displayed to the user and the user can input the values from upload.html

Conclusion:

Thus we have trained and tested the model using the model i.e Decision tree with least root mean square error and trained it on ibm cloud and also deployed it and integrated Flask with the scoring end point and did the same locally by saving the file using pickle and designed the user interface with html.

References:

- https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=functions-watson-machine-learning-overview
- https://www.upgrad.com/blog/top-python-libraries-for-machine-learning/

- https://www.geeksforgeeks.org/
- https://towardsdatascience.com/a-practical-introduction-to-9-regression-algorithms-389057f86eb9
- https://bookdown.org/tpinto_home/Regression-and-Classification/k-nearest-neighbours-regression.html