Baba Siddharth

Application Development Associate – Machine Learning Engineer

Address Bangalore,Karnataka,560078

Phone 9742784704

E-mail babagunner1581@gmail.com

LinkedIN linkedin.com/in/baba-siddharth-25b39a160

Github ***https://github.com/baba1581***

Tableau https://public.tableau.com/app/profile/baba.siddharth

Aspiring Data Scientist with knowledge on gathering, cleaning and organizing data for use by technical and non-technical personnel. Advanced understanding of statistical, algebraic and other analytical techniques. Highly organized, motivated and diligent with significant background in Machine Learning, Python and SQL.Willing to learn new technologies that comes up.

Work History

Accenture Services Pvt Ltd, Bangalore, India **2021-01 - Current**

Project: BioCentury

Project Role: Machine Learning Engineer

Technologies: Python (V 3+), SQL (MySql), Data Science, Machine Learning

- Handled missing values by applying classifier algorithms ,implemented and applied **Label** and **One-Hot encoding** for categorical data, **Outlier** detection using Inter-Quartile Range, Percentile, and Z-score.
- Used **confusion matrix** and handled imbalanced classification data to build a better model.
- Used **feature scaling** for non-tree based models to improve the accuracy by 30%.
- Implemented **skewness treatments** such as Box-Cox, Square Root, Reciprocal, and Log Transform that helped to **decrease** the **skewness** by more than **98%.**
- Implemented **Natural Language Processing** techniques for text preprocessing such as stop-word removal, stemming, lemmatization, Bag of Words, TF-IDF, Word2Vec, and word embeddings.
- Used hyperparameter optimization techniques like **GridSearchCV** and **RandomizedSearchCV** to improve the model accuracy.
- Used regression algorithms like : **Random Forest ,XG Boost ,Ridge ,Lasso.**
- Used classification algorithms like : **Random Forest ,Gradient Boost ,XG Boost, SVM ,Logistic Regression** and unsupervised clustering algorithms **like K-Means ,Hierarchical and DBSCAN.**

Saarthi.ai, Bengalore,India **2019-01 -2019-07**

Project Role: NLP Engineer Intern

- Performed data scrapping using **BeautifulSoup bs4**.
- Used various data preprocessing techniues like handling missing data ,data transformation ,dimensionality reduction techniques like **PCA.**
- Used text preprocessing techniques like TF-IDF ,BOW ,Word2Vec and trained the model on various machine

learning algorithms to find that **XG Boost** gave the best accuracy with 98%.

Education

2016-08 - 2020-06 **Bachelor of Engineering: Information Science And Engineering**
Dayananda Sagar College of Engineering - Bengaluru, KA

2021-03 - 2021-12 **Post Graduation – Data Science**
Purdue University

Projects

❖ Movie Recommendation System with Sentiment Analysis: GitHub

- **Content-Based** Recommender System recommends movies similar to the movie user likes and analyses the sentiments on the reviews given by the user for that movie.
- Extracted data from the IMDB dataset from kaggle, Wikipedia, and TMDB website using tmdb **api key** and **beautifulsoup4**.
- Performed EDA, data preprocessing etc to convert data to the required format.
- Analyzed review's sentiments of the movies using **TF-IDF vectorizer** and **Naïve Bayes** classifier and got 99.2**%** model accuracy and calculated similarity using **Cosine Similarity**.
- Deployed it on Heroku using Flask.
- URL : https://movie-recommendation-baba.herokuapp.com/

❖ Flight Fare Prediction: GitHub

- Predicticting the price of flights across various locations in India to help you book the most relevant flight for your journey.
- Used kaggle dataset ,performed EDA, feature engineering, handled categorical and data time type data.
- Selected important features using **ExtraTree Regressor** and **correlation.**
- Fitted the model using **RandomForest Regressor**.
- Performed **hyperparameter tuning** using RandomizedSearch CV to increase the accuracy.
- Deployed it on Heroku using Flask.
- URL : https://flight-price-prediction-baba.herokuapp.com/

Skills

- Data Science
- Machine learning
- Python
- SQL
- Deep Learning
- Natural Language Processing
- Exploratory Data Analysis
- Data Visualization(Tableau)
- Web Scraping
- Statistics