Hrushikesh Sahu

Data Science Trainee at AlmaBetter | Sambalpur

hrushikeshsahu19@gmail.com | 7848866575 | linkedin.com/in/hk-sahu | github.com/hrushikeshsahu19

PROJECTS

Meru taxi trip time prediction
AlmaBetter Verified Project

12/2020 - 01/2021 *Bangalore*

Tags: Regression, XGBoost, Gradient boosting machine, MSE, R-square, Decision tree, VIF, homoscedasticity, multicollinearity, Gridsearch CV, feature engineering, Lasso, Ridge, Pearson correlation

- Built a **regression model** using **GBM, Decision tree regressor**, and **XGBoost models** to **predict taxi trip time** in Delhi for a time period of six months.
- Used a **folium graph** to visualize the pick-up and drop-off locations and used **heatmaps**, which were essential for **EDA.**
- Applied **feature engineering** to obtain new features such as distance, speed, peak hours, busiest days and used **Pearson correlation, VIF values** to avoid **multicollinearity in Linear Regression.**
- Applied **Lasso** and **Ridge** regularisation for optimizing the fit of the model and used **Gridsearch CV** for **hyperparameter tuning**, which resulted in **R- square** score of **0.71** on the test dataset.

Marketing Campaign Effectiveness Prediction
AlmaBetter Verified Project

01/2021 - 02/2021 *Bangalore*

Tags: PCA, Anomaly Detection, Feature engineering, Imbalanced dataset, Oversampling, Shapley Additive exPlanations (SHAP), Isolation Forest, SMOTE, Marketing Campaigns

- Developed a **stacked model using Logistic meta-classifier** on-base classifiers such as **XGBoost**, **SVM**, and **Random Forests** to predict whether a customer will start a **Fixed-Deposit** as a result of a marketing campaign. Obtained the **ROC-AUC score** of **91%** on the test data.
- Treated **multivariate outliers** using **Isolation Forest** and applied **SMOTE** boosting on normalized data to mitigate the problem of class **imbalance.**
- Used **SHAP values** to determine the most important features contributing to purchase such as the **number of calls made during the campaign**, **bank balance, personal loan, housing loan** etc.

News popularity prediction on social media
Almabetter Verified Project

02/2021 - 03/2021 *Bangalore*

Tags: Regression, NLP, spacy, PCA, Random forest, XGBoost, GBM, MSE, R-square, dimensionality reduction, TFIDF vectorizer, tokenization,nltk, feature engineering, Bias-Variance

- Built a **regression model** to predict the **popularity of news** on social media platforms such as **Facebook, LinkedIn, and Google Plus**.
- Used **tokenization, lemmatization**, and pos-tagging and leveraged the concepts of **SpaCy** library to carry out **text processing** on the given dataset.
- Applied **TFIDF vectorizer** along with **PCA** to reduce the complexity of the dataset and used **Gradient Boosting Machine, Random forest**, and **XGBoost** models to come up with the best working model on the test dataset for each of the three platforms respectively.
- Carried out **bias-variance tradeoff** analysis to optimize the fit of the model and used **Gridsearch CV** for **hyperparameter** tuning, which helped in achieving **R-square score of 0.72, 0.76, 0.78** respectively for each of the social media platforms.

TECH STACK

Languages
Python, C,JAVA

ML Frameworks
Scikit-learn, Keras, Pandas, Numpy, Seasons, Matplotlib, spaCy, Keras, Pandas, OpenCV, NLTK, Plotly

Platforms
Jupyter Notebook, Google Colab, Spyder, MS Office

Databases
SQL, Oracle

EDUCATION

B.TECH In Computer Science and Engineering
Indira Gandhi Institute Of Technology, Sarang

2021

8.55/10.0

XII-Higher Secondary
Yuvodya Junior College, Balangir

2016

77.16/100

X-Secondary
Govt. High School, Kuhibahal

2014

75.33/100

RELEVANT COURSESWORKS

Machine Learning(AppliedAI) (2020)
KNN, SVM, Bagging, Random Forest, Naive Bayes, Boosting, GBDT, Xgboost, K-Means, PCA, LDA, NLP

Statistics for Data Science(Udemy) (2019)
Probability distribution, confidence interval, Hypothesis Testing, central limit theorem,Co-relation , Regression

ACHIEVEMENTS

Awarded Scholarship for Higher Secondary Education by State Government, 2016

1st in District Level Cricket Competition, 2013

INTERESTS

Music | Cricket | Swimming