Rishi Kumar

Chennai
7708520314
rishikumar4197@gmail.com
github.com/Rishikumar04
linkedin.com/in/rishi-kumar4197

Summary

Aspiring data scientist with hands on experience in executing data driven solutions to increase efficiency, accuracy and utility of internal data processing. Experienced at creating machine learning model using various machine learning algorithms with in-depth understanding.

Skills

- Proficiency in Python Language (Numpy, Pandas, Scikit-learn).
- In-depth understanding of Predictive Analysis using Machine Learning algorithms.
  - Modelling:
    - Supervised Learning : Linear and Logistics Regression, tree based algorithms, Support Vector Machines, Naive Bayes, K-Nearest Neighbor and Ensemble Method.
    - Unsupervised Learning : K-means clustering, Hierarchical clustering and Principal Component Analysis(PCA).
    - Time Series: ARIMA, SARIMA, FBProphet.
- Created Predictive model for Hospital cost estimation using Streamlit framework.
- Data Visualization using Tableau, Seaborn, matplotlib.
- Expertise in Data mining , Data Cleaning and modelling
- Published article about Linear Regression in Medium.
- Basic understanding of Structured Query Language (Postgres).

Projects

- **News Category Prediction.**
  - **Overview:**

    Predict the genre or category a piece of news based upon the storyline.

  - **Techniques:**
    - Predicting news category by using storyline as input variable.
    - Encoded Storyline using various text extraction techniques like CountVectorizer, TfidfVectorizer, HashingVectorizer.
    - Categorized news from text extracted data with machine learning algorithm like Logistic Regression, Random Forest, XG Boost, CAT Boost.
  - **Reference**:
    - https://github.com/Rishikumar04/Practise-Problems/blob/main/Machine%20Hack/02_Predict_News_Category_Machine_Hack.ipynb
- **Bangalore House Price Prediction**
  - **Overview:**
    - Estimation of house cost considering locality, size, availability with lots of hands-on preprocessing data.
  - **Techniques:**
    - Explored various relations in dataset using visualization.
    - Implemented mean encoding techniques for converting categorical variables.
    - Preprocessed Total sqft column, which has impurities in data like in various units and brought them all to uniform unit total square feet.
    - Overall house cost is predicted using various machine learning algorithms, among those XG boost gave better accuracy.
  - **Reference:**
    - https://github.com/Rishikumar04/Practise-Problems/blob/main/Machine%20Hack/05_Bangalore%20House%20Price_Machine_Hack.ipynb

- **Predicting Data Scientist Salary**
  - **Overview:**
    - Annual salary range estimation for the Data Scientist from the YOE, job designation and location.
  - **Techniques:**
    - Predicting Data Scientist salary to which category they belong.
    - Implemented preprocessing techniques and encoding using Label Encoder.
    - By using XGBoost algorithm predictive model was created.
  - **Reference:**
    - https://github.com/Rishikumar04/Practise-Problems/blob/main/Machine%20Hack/03_Data%20Scientist%20Salary%20Prediction_Machine%20Hack.ipynb

- **Hospital cost estimation**
  - **Overview:**
    - Total cost of hospitalization is estimated by several categories such as disease, mode of arrival etc..
  - **Techniques:**
    - Streamlit library is used to create the User Interactive local app. The final outcome result was compared by using different machine learning algorithms like Linear based models, tree based models.
    - A user can give manual inputs to predict the cost of the hospital, which is based on the model trained on the dataset.
  - **Reference:**
    - https://github.com/Rishikumar04/Stream_lit

- **Employee Attrition prediction**
  - **Overview:**
    - Predicting the probability of employee attrition by considering various factors like Environment Satisfaction, Work Life Balance, Salary Hike etc...
  - **Techniques:**
    - Explored Kaggle's employee attrition data and visualized several business insights.
    - Pandas library's get dummy method was used for converting categorical data into numeric data type
    - Standard Scaler technique was applied to form the data into uniform scaling units.
    - The employee attrition predictive model was designed by CatBoost classifier.
  - **Reference:**
    - **https://www.kaggle.com/rishikumar97/rishi-employee-attrition-prediction**

- **Health Care Analytics**
  - **Overview:**
    - Patient's length of stay in hospital is predicted by taking account of several factors.
  - **Techniques:**
    - Various business insights were derived from the Health Care Analytics data.
    - Category Encoder library's Leave one out Encoder was applied to encode categorical data.
    - StackingCVClassifer algorithm from mlxtend library was used to derive the final predictive model.
  - **Reference:**
    - **https://github.com/Rishikumar04/Health-Care-Analytics--Hackathon**

Training & Certifications

- Vice President - Public Relations of Tirunelveli Toastmaster Club.
- Completed 2020 Complete Python Bootcamp from Zero to Hero by Jose Portilla in Udemy
- Data Science Certification from Inceptez Technologies.

Education

B.E Mechanical Engineer 2014 - 2018

Karpagam institute of technology

CGPA-7.23

Higher Secondary. 2013 - 2014

St. Xavier's Matric. Hr. Sec. School,Tirunelveli

Percentage-89.8%

Secondary 2011 - 2012

St. Xavier's Matric. Hr. Sec. School, Tirunelveli.

Percentage-90. 2%