Sarvesh Mishra

Kaggle ● Github ● LinkeDln

✉ sarveshmishra.bhu20@gmail.com
📞 +91 8957883380

Educational qualification

Bachelor of Science (BSc), Banaras Hindu University

2020–2023
CGPA – 7.6/10

Data Science Course & Certification

iNeuron Full Stack Data Science Masters

06/2023 - 02/2024

Technical skills

- **Machine Learning** : Linear, Classification, Predictive unsupervised learning
- **Big Data :** Hadoop, Spark
- **Data analysis & visualization tools:** Advance Microsoft Excel, Power BI, ETL, Google Data studio
- **Language and Frameworks** : Python, R, Scikit-learn, Flask
- **Database:** SQL, Mysql, Snowflake
- **Libraries** : Numpy , Pandas, scikit-learn
- **Natural Language Processing (NLP)** : NLTK, Word2Vec
- Statistics, Linear algebra, Probability
- **Others:** Time Series, AWS

Projects

1 Kaggle Mohs Hardness Competition

View Project

- Kaggle hosted a competition where various attributes of minerals were provided, and participants had to determine the hardness of minerals using machine learning algorithms.
- Criteria of evaluation was median absolute error
- Used XGBoost, RandomForest Regressor, Gradient boosting algo and ensemble learning to predict the hardness
- Applied GridsearchCV for best parameters
- Performed Data cleaning, data scaling and addressed overfitting of data
- As a result, concluded the competition with a rank of **734th out of 1500+** competitors

2. Laptop Price Predictor Using Regression Model

View Project

- **Objective:** The objective was to develop a regression predictive model to predict the absolute price of laptops based on various input variables. The project utilized Python, machine learning models, Pandas for data manipulation, regression techniques, and statistical analysis.
- Employed Pandas for data cleaning tasks, such as imputation of missing values or removal of irrelevant information. Engaged in feature engineering to create new relevant features or transform existing ones to improve the model's performance
- Evaluated the performance of the regression model using relevant metrics such as **Mean Absolute Error** (MAE) and Root Mean Squared Error (RMSE).
- Achieved a predictive accuracy score of **86%**, indicating the model's ability to estimate laptop prices.

3. Swiggy Customer order details analysis of the 9 different cities

View Project

- **Objective:** The primary goal was to derive meaningful insights, both in terms of numbers and graphical representations, to facilitate informed business decisions and improvements.
- Cleaned and update the dataset as business requirements
- In total, there were **8,680** orders across nine cities, with the highest percentage **15.50% originating from Kolkata**.
- Revenue analysis reveals that Mumbai contributes to **17%** of the total revenue, which amounts to **0.56 million IN**R.
- Notably, the maximum revenue per order is from Mumbai at **393 INR**, while the minimum is from Surat at **270 INR.**
- The average delivery time across all cities is **54 minutes**. The slowest delivery time is in Kolkata, taking **68 minutes**, whereas the fastest is in Mumbai, with an impressive **48 minutes**.
- Built the graph for understanding the complex statements.
- **Tools used :** Microsoft Excel, Power BI, SQL, Snowflake

4. Airline Passenger's forecasting using Time Series Models

- **Objective:** Build a time series model for forecasting customer numbers over the next **10-20** years.
- Implement Dickey-Fuller test to assess data stationarity.
- Utilize differencing and rolling methods to convert the data to a stationary state.
- Apply the **ARIMA** and **fbprophet** model for accurate and effective forecasting. Concentrate on time series analysis and prediction.
- Provide insights into future customer trends.

5. Diabetes Data Analysis and Prediction

- **Objective:** The primary objective was to analyze and visualize a dataset related to diabetes, aiming to identify the factors that contribute to the occurrence of diabetes with analytical report.
- Conducted thorough exploration of the dataset to understand its structure and characteristics. Utilized statistical techniques and graphical methods to summarize and interpret the main features of the data. Visualized relationships between different independent variables and the target variable (Outcome) to identify potential patterns and correlations.
- Implemented a advance machine learning model using regression techniques. Chose an appropriate regression algorithm based on the nature of the problem and dataset.

Achievements

Achieved Kaggle 2x Expert status through consistent participation and high-performance rankings in data science competitions. Demonstrated expertise in applying advanced machine learning techniques and collaborating with a global community of data scientists

4 Star SQL on HackerRank

About me:

With a solid foundation in mathematics, physics, and geology acquired during my BSc studies at Banaras Hindu University, I have developed a strong analytical mindset and a keen interest in problem-solving. Now, I am enthusiastically transitioning into the field of data science, driven by my passion to tackle complex data-related challenges and harness the power of data-driven insights to make a meaningful impact and contribution.