Sumit Kumar
Data Scientist

✉ **sumitme752@gmail.com**   💬 (+91) 8553282301   📍 Bangalore, India

☰ Career Summary

Motivated professional with 4 years of experience in Data Science and Analytics (total exp 6 years) . Experienced at interpreting and analyzing data to drive solutions, and furnish insights and business intelligence needed to guide decisions for a business problem. Passionate about building models that fix issues. Relevant skills include machine learning, problem-solving, programming and creative thinking. Also, I have experience in training & guiding individuals on ML and related topics.

•⌁ Skills

- **Statistical Analysis –** Exploratory Data Analysis, Inferential Statistics, Hypothesis Testing
- **Machine Learning & AI** –

    Algorithms: Decision Tree, Random Forest, XG Boost, Linear Regression, Logistic Regression, SVM, KNN

    NLP: Text preprocessing, Deep Learning - ANN , RNN ,CNN
- **Programming Languages & Databases** - Python , SQL
- **Frameworks** - NumPy, Pandas, Sci-kit learn, SciPy
- **Tools/IDE** - VS Code, Jupyter Notebook , Excel , Git
- **Operating Syste**- Windows , Linux , MacOS
- **Web Development** - HTML, CSS, ReactJs

💼 Work Experience

1. Pravartan Technologies – Data Scientist (Mar 2023 – Present)

Projects

1. Data Quality Analysis and Dashboarding

Goal : This project focuses on assessing and visualizing the completeness and accuracy of data using a combination of tools and technologies, including Excel, Python, Power BI, and Snowflake. The primary goal is to ensure the integrity of data and provide actionable insights through comprehensive dashboards.

Roles & Responsibilities

- Data Extraction: Connect to Snowflake and extract relevant datasets.
- Data Analysis: Using excel perform preliminary data checks, cleaning, and basic completeness analysis. Conduct advanced data analysis to identify incorrect data and automate quality checks.
- Dashboard Development: Used Power BI to create dynamic dashboards that provide visual insights into data quality metrics.
- Review and Reporting: Compile findings into comprehensive reports and suggest actionable recommendations for data quality improvement.

2. Automated Form Filling Using Data Preparation and Analysis

Goal : This project demonstrates a proof of concept (POC) for automating the process of filling blank forms with valid answers. Focuses on data preparation and analysis, leveraging Azure Computer Vision to convert PDF forms into JSON format for further processing. This will significantly reduce manual effort and increase accuracy.

Roles & Responsibilities

- Data Preparation: Collect and clean data required for filling forms, ensuring it is structured and ready for analysis.

- Data Analysis: Conducting thorough analysis to validate the data, ensuring accuracy and completeness for form entries and documenting the data analysis process and validation results.
- PDF to JSON Conversion: Used Azure Computer Vision to convert PDF forms into JSON format, enabling structured data extraction.
- Automated Form Filling: Develop scripts to automate the process of populating blank forms with valid answers based on the prepared and analyzed data.

2. Skoruz Technologies - Data Scientist (Jan 2022 - Sept 2022)

Projects

1. Goal - To help J&J with "Prediction of Leadtime" and "Choosing the best lead time prediction value" among the different lead time predictions available.

Description - Leadtime is no. of delivery days required to deliver an incoming order. The prediction we provided is called "d2" prediction; similarly, J&J has other mechanisms to make the predictions. So providing them with solutions and helping them with choosing the best among the available prediction values. The project has four components - data extraction & pre-processing, D2 training, method selection training, and live prediction.

Roles & Responsibilities

- EDA - Treatment of missing values, Encoding of categorical features, Capping & flooring for outliers
- Decision trees, Random forest, Logistic regression, and XGB are used for model building.
- The model with the best performance metrics is XGB with an accuracy of 89% and precision of 72%.

3. Flatworld Solutions - ML Engineer (May 2019 - July 2021)

Projects

1. Goal - Create a proactive customer retention strategy for a telecom company. To identify which customers would churn for a telecom company based on their usage behavior. This is a Customer Retention Strategy. A predictive model is created using classification algorithms and based on the prediction, a strategy can be made to target the profitable customers and retain them.

Roles & Responsibilities

- Exploratory data analysis to get better insights by treating missing values, outliers, and null values.
- Encoding of categorical features and analysis of correlations between numeric variables.
- Decision trees, Random forest, and Logistic regression are used for model building.
- To improve model performance, hyperparameter tuning using grid-search cv for Decision Tree and Random Forest.
- The model with the best performance metrics is Random Forest, and was chosen to create the strategy.

4. Rama Engicons (India) Pvt. Ltd - Design engineer (October 2016 - April 2019)

- Modelling and Detailing of Sheet Metal Components-The objective of the project was to convert the 2D drawings to 3D sheet metal models that could be directly fed for manufacturing.
- Soundproof Enclosure- This project was about sound and leak-proof enclosures.

✏️ Education

2012-2016	2021
New Horizon College of Engineering (VTU University)	**Data Science and AI Certificate by IBM (Learnbay)**
B.tech	

🔖 Language

- English , Hindi , Kanada

🔖 Interest

- Mixed Martial Arts , Trekking , Cricket