SHUBHAM DUBEY

(973) 979-4949 | sd858@njit.edu | www.linkedin.com/in/dubshubham | https://github.com/shbhamdbey

EDUCATION

New Jersey Institute of Technology, *MS in Computer Science, NJ*   3.50 GPA   Aug 2019 - May 2021
D.Y Patil College of Engineering, *BE in Information Technology, India*   3.48 GPA   Jul 2014 - Jun 2018

PROFESSIONAL EXPERIENCE

Student Assistant (Data Quality Analyst), NJIT, *Newark,* **NJ**   **Sept 2020 – Present**
- Worked on extensive data validation using **SQL** and maintain back-end testing reports using Excel extract and increasing data quality assurance via 12%, communicating results using **HLOOKUP** and **PIVOT** dashboards.
- Optimized the already stored **cursor and trigger** to enhance the performance with the **ERP** banner data mapping and identifying potential area where exploratory data analysis can be used to increase efficiency using daily scrum
- Performed data cleaning of source data stored in **SQL server** and maintaining deduplication chain cleanup for the data centered products which helped in handling critical business issues and increased efficiency by 20%

Capstone Fellow Supply Chain Analyst, Pfizer, *Newark, NJ*   **July 2020 – Jan 2021**
- Extracted and transformed geographical information and developed a **Risk analysis model** for prioritizing with response strategies to improve operational efficiency in logistics by 5%
- Assisting with **Python scripts** and pipelines for spatial data analysis, providing optimal storage in **MongoDB Clusters** which helped in reducing waiting time for analysis of the reports by 20%
- Communicated **recommendations** of Risk Data Model using Tableau reporting dashboards having interactive maps and charts to improve latency by 49% to 63%
- Worked with Pfizer global supply chain on data driven approach to analyze external entities impacting clinical trials sites, which helped in generating actionable insights and to increase proactiveness
- Designed and Created computational and **statistical modeling** applications for geospatial time series data analysis and segregating external event recognition Using Python and **Tableau**.

PERSONAL PROJECTS

Sales Data Reporting:
- Summarized business reports for the sales team over time series data, discovering key areas to save spending by over 23,000$ per year utilizing the power of **Tableau** dashboards.

Quality Prediction:
- Accelerated statistical analysis over wine quality data in **AWS S3** using **PySpark** and AWS EC2 parallel cloud instances with providing centralized platform free deployment using **Dockers** containers.
- Implemented bootstrapped approach using random forest which lead to **F-score** to 78% predictions.

Image recognition pipeline:
- Utilized the power of AWS ML **API** image rekognition with confidence higher than 90%, mapped EC2 instances using **SQS queue** to perform parallel label and text prediction.

Easy School Admit Predictions:
- Automating the school application system where factor impacting school admit and reject where analyzed using **logistic regression** with 80% accuracy and deploying it with **Python Flask**

Prediction of US illegal immigrants:
- Performed trend and seasonality analysis over time series data using **ARIMA model**, forecasting future illegal infiltrations in USA, identifying the most common point of entry preventing infiltration by 8%
- Visualized and reported the most frequently ports and channels which **Tableau dashboard** and **Excel**.

SKILL

Analytics and Databases: MySQL, SQL Server, Tableau, Oracle, Alteryx, Python, R, ETL, Aws ML, C++, Data mining, Big Data, Business Intelligence.
Tool: PowerPoint, Word, Excel, Pivot tables, VLOOKUP, JIRA, Putty
Skills: Time Series Forecasting, Risk analysis, Data Management, Data Visualization, Agile, Data Modelling, Decision Analysis, Statistical Modelling, A/B Testing, Marketing Analytics, Trend Analysis, Predictive Analysis