ADITYA DUBEY

Chicago, IL | adubey6@hawk.iit.edu | (312)918-8779 | LinkedIn | GitHub

EDUCATION

Master of Information Technology and Management DEC 2020
Illinois Institute of Technology, Chicago, IL, USA

- **Courses:** Data Analytics, Data Mining, Data Warehousing, Database Management, Project Management

Bachelor of Engineering in Computer Science JUNE 2018
Shri Govindram Seksaria Institute of Technology and Science, Indore, India

TECHNICAL SKILLS

- **Data Science:** Classification, Linear Regression, Multiple Regression, Logistic Regression, Naïve-Bayes, KNN, Time series analysis, Random forest, Decision tree. Clustering and modeling- association rules, Outliers detect
- **ML Libraries**: NumPy, SciPy, Pandas, Scikit-learn
- **Languages:** R, Python, Java, C, C++, HTML, CSS, JavaScript
- **Reporting & Visualization:** MS Power BI, Tableau, Excel.
- **Tools**: R studio, Jupyter Notebook, Power BI, Tableau, SQL server management tool, SQL server data tool, Microsoft Office, Git, Azure DevOps, Microsoft excel.
- **Database:** MS Access, Oracle SQL Developer

WORK EXPERIENCE

Data Analyst, *Avocare Biotech, India* Jun2018 – July 2019

I worked on a **research project** in a startup- we were developing a rare medicinal mushroom in a laboratory, which helps **treat cancer**. I provided temp and humidity predictions for the plant to sustain life in a laboratory for each stage by using Predictive modelling.

- Analyzed1.3 million rows to check the effectiveness of the drug & Provided visualization.
- Forecasted the effective temperature for the drug using the decision tree.
- Analyzed the data and generated the clusters with similar features in the drug using K-means algorithm.
- Figured out 3 important patterns by time series analysis with an accuracy of 89%.
- Solved the Problem of human interaction in the lab by creating the IOT device for lab automation, which increased the sustainability of drug by 33%.
- Increase productivity by 40% of the lab by providing an idea of different slot experiment with different temp, result from data science predictive modelling.

ACADEMIC PROJECTS

Rossman Sales Prediction (R, Predictive modelling, MS Excel) - (GitHub- Report, code)

- Utilized 1Million Rows with 7 attributes to predict the pattern in Sales for Business growth.
- Performed data Preprocessing (created N-1 variables for categorical variables).
- Implemented 5 models (Random forest, Decision tree, Naïve Bayes, K nearest neighbor, MLR)
- Compared all models and found Best Accuracy of 89.5% achieved using Random forest regressor.

Earthquake Prediction 2021-(Big Data Pipeline PySpark, MLlib, MongoDB, Bokeh, Flask, Jupyter) -(GitHub)

- Implemented Machine learning Models & created predicted Dataset for earthquake 2021.
- Created ETL Pipeline for large earthquake data set & performed Data Preprocessing (PySpark, MongoDB).
- Built visualization and interactive dashboard using Bokeh library and Tableau (heatmap, geo map, Magnitude plot).
- Created the Pyspark ETL Scripts & Virtual Dashboard Server using Flask.

Crop Yield Prediction (Regression, Python) (GitHub- Report, code)

- Forecasted the 5-year crop yield report by utilizing 2millions rows with 17 features.
- Implemented memory optimization by 72% & performed heat map analysis.
- Optimized classification models to Improve Prediction (feature selection, cross-validation | 7% improvement)
- Compared all models and build Yield prediction graph by Decision tree with MAE value of 70%.

COVID-19 Analysis (PowerBI, Tableau) (GitHub- Dashboards)

- Created interactive dashboard (geo-map) & performed analysis on Covid-19 data .
- Analyzed Covid-19 data & identified patterns in confirmed, recovered & new cases according to countries.

CERTIFICATION AND PUBLICATION

- **Tableau Analyst**, Tableau Software, 2020
- **Machine learning**, Stanford Online, 2018
- **Publication** – "*An Effective TCP's Congestion Control*", Springer International, 2019