

ADITYA DUBEY

📍 Chicago, IL | ✉️ adubey6@hawk.iit.edu | ☎️ (312)918-8779 | [in 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, regressions, outlier detections.
- **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, Jupiter Notebook, Power BI, SQL Developer, 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 **research project** in a startup- we were developing a rare medicinal mushroom in a laboratory, which helps **treat cancer**. I was providing temp and humidity predictions for the plant to sustain life in laboratory for each stage by using Predictive modelling.

- The role was focused on data extraction, Transformation and analysis and Prediction modelling.
- Provided visualization of data to check the effectiveness of the drug.
- Find the effective temperature range for the drug using the decision tree.
- Analyzed the data and generated the clusters with similar features in the drug using K-means algorithm.
- Provided all possible Temp & Humidity Prediction Report for each stage with accuracy of 91percent.
- Solved the Problem of human interaction in lab by creating the IOT device for lab automation which increased the sustainability of drug.
- Increase the productivity of the lab by providing idea of different slot experiment with different temp, as a result of data science predictive modelling.

ACADEMIC PROJECTS

Rossmann Sales Prediction (Python, R, Predictive modelling, MS Excel)

- Performed data Preprocessing and built predictive models (Linear Regression, Random Forest, Decision Tree) and Neural Network to predict the future sales of Store.
- Accuracy of 89% achieved using Random forest regressor.
- Provide the Business decision on promos using Hypothesis testing.
- Preprocessing of data- extracting, cleaning, transforming, analyzing, and visualizing massive amounts of data.

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

- Implement 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 (Classification algorithms, R)

- Performed data analysis and predicted next 5-year crop yield using classification algorithm like KNN, Naïve Bayes Logistic Regression, XGboost with 91% accuracy.

COVID-19 Analysis (PowerBI, Tableau)

- Created interactive dashboard (geo map) & performed analysis on Covid-19 data .

CERTIFICATION AND PUBLICATION

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