

Abhimanyu Gangula

Tampa, Florida | gangulaabhimanyu@hotmail.com | [LinkedIn](#) | [Tableau](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Data Scientist with 2+ years' experience as a Data Analyst. Well-versed with statistical data modeling and skilled at deploying appropriate ML models to solve business problems. Adept at translating actionable data insights into business value. Proficient in **Python, SQL, Tableau**, Machine Learning Algorithms and libraries.

ACADEMIC EDUCATION

University of South Florida

Jan'21 - Dec'22

Master's in Business Analytics and Information Systems

University of Petroleum and Energy Studies

July'13 - May'17

Bachelor's in Applied Petroleum Engineering with Specialization in Upstream

TECHNICAL SKILLS

Programming: Python, R, C#, SQL

Big Data: Apache Spark, Apache Hadoop, MapReduce, Hive

Statistical Analysis: Regression Analysis, A/B Testing, Inferential and Descriptive Statistics, Survival Analysis

Machine Learning: Time-series forecasting, Neural Networks, Natural Language Processing (NLP)

Libraries: Pandas, NumPy, SciPy, Scikit-Learn, TensorFlow, PyTorch, Keras, Seaborn

Visualization: Tableau, Power BI, Excel, Visio (UML)

WORK EXPERIENCE

Data Science Intern – **iQuest Solutions**, Texas

Aug'22 – Present

Time Series Forecasting

- Working on building a **time-series model** in the Aviation domain to predict the busiest airlines and recommend better routes for a given season using multi-variate analysis in Python.
- Gathered, mined and analyzed data from authorized sites to identify patterns for time-series data
- Implemented forecasting algorithms such as **ARIMA, SARIMA** and neural networks (**RNN, LTSM**) using time series forecasting libraries, such as Prophet, Statsmodels, and scikit-learn.
- Evaluated the performance of time series forecasting models using metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

Data Science Intern – **FCCI Insurance**, Florida

June'22-Aug'22

Predictive Modelling

- Developed **underwriting** and **claims models** by creating predictive features, utilizing external data, and applying statistical and machine learning techniques as **Logistic Regression, Random Forest, Extreme Gradient Boosting**.
- Gathered a good knowledge of Insurance domain and industry specific terminology
- Successfully applied predictive models to real-world FCCI insurance problems, such as pricing, underwriting, claims management and customer segmentation.
- Experience with evaluating model performance using metrics such as Accuracy, Precision and AUC-ROC.

Data Analyst - **Tech Mahindra Ltd**, India

May'17-Jul'19

- Mined, analyzed, and manipulated databases with 20 years of data using **SQL** and **Excel** for diverse business requirements to generate data reports and key business operations strategies.
 - Designed interactive, data-driven dashboards and scorecards using **Tableau** and business intelligence tools to monitor real-time data.
-

RELEVANT PROJECTS

- **Financial Risk Estimation using Monte Carlo Simulations** Nov'22
- Assessed risk of an investment portfolio using Monte Carlo simulation.
 - Scraped stock data from Yahoo using **REST API calls** and loaded them onto a PySpark dataframe.
 - Trained Considered market factors **GSPC value, IXIC value, return of crude oil, return of treasury bonds** to estimate the **risk variable (VAR)** over the next two weeks with a 95% confidence interval.
 - Evaluated the results of the simulation using a **back-testing** method.
- Tools & Technologies used:* PySpark, HDFS, Jupyter Notebook
- **Sentiment Analysis of product reviews of an e-Commerce platform** May'22
- Built a classifier to predict the reviews of products from Flipkart (e-Commerce) into their respective classes.
 - Loaded the reviews onto a spark session and pre-processed the dataset.
 - Processed the reviews with *NLP* techniques like *Tokenization*, stop words removal and built a *tf-idf vectorizer* as input to machine learning models (Random Forest Classifier, Linear SVC and Logistic Regression).
 - Built the classification model and evaluated the models to dish out the best predictive algorithm.
- Tools & Technologies used:* Databricks, PySpark and Python
- **Influence of socioeconomic factors on Incarcerated population** Apr'22
- Built comprehensive data models such as Poisson distribution, MLE and Quasi-Poisson Distribution in *R studio* to analyze the influence of socio-economic factors on prison population.
 - Collected data from various sources such as *prisonpolicy.org*, *data.census.gov* and multiple websites and pre-processed data along with feature extraction to make the dataset ready for analysis.
 - Examined the correlation effects and skewness in the data with plots, graphs and data visualization techniques.
 - Presented the marginal effects of each variable and made actionable recommendations to mitigate the influences.
- Tools & Technologies used:* Tableau, R and R studio
- **EDA in Film Industry for a successful movie studio** Nov'21
- Webscraped data from websites such as *imdb*, *moviefone*, *boxofficemojo* etc., to extract data using *BeautifulSoup*.
 - Loaded data onto a dataframe using pandas and cleaned/pre-processed the data.
 - Visualized the data in Tableau to answer questions such as :
 - a. How much should a studio spend to make a successful movie
 - b. What genres and age groups should the studio focus on
 - c. What time of the year should a studio slate their releases
 - Made recommendations and presented results for running a successful movie studio.
- Tools & Technologies used:* Tableau, Python and Jupyter Notebook
- **Predicting Drill bit wear and Reservoir Formation using ML techniques** June'21
- Analyzed data (drilling & logging) provided by Equinor on Volve field (North Sea) made public in 2018.
 - Problem statement is two-fold: To predict drill bit wear (Regression) and Reservoir type (Classification).
 - Developed and fine-tuned the ML algorithms for Formation Classification using K-Neighbors Classification(12 ¼) and Gradient Boost Classifier(8 ½) with an accuracies over 92% and 76% respectively.
 - Compared different ML algorithms like Decision Tree, Random Forest and Ada Boost with accuracies over 95 % to predict drill bit wear with corresponding sections (26", 17 ½, 12 ¼, 8 ½) accordingly.
- Tools & Technologies used:* Python, Jupyter Notebook, Tableau, Pandas, Numpy and Seaborn
-