SYED SIRAJ

syed.siraj19151421@gmail.com | +91 7730942589
LinkedIn: https://www.linkedin.com/in/siraj-syed-456125139/
Academic Portfolio: https://eportfolio.mygreatlearning.com/syed-siraj

PROFILE:

Passionate to work with maximum potential in challenging and innovative environment, with an opportunity of working with a diverse group of people and enhance my professional skills with learning and experience.

PROFILE SUMMARY:

- Data professional with 3 years of hands-on experience on analysing and architecting and implementing robust data pipelines in Snowflake, Airflow, and DBT.
- Having certifications in Microsoft Power BI and Azure Data Fundamentals, showcasing my expertise in cloud technologies. Proficient in utilizing BI tools such as Qlik Sense and Tableau to extract meaningful insights for business stakeholders.
- Have completed courses in Big Data Analytics and am currently pursuing a Post Graduation Program in Data Science and Business Analytics.
- Skilled in applying various machine learning models and algorithms for predictive analytics & Forecasting in Python and PySpark.
- My experience extends to data preprocessing, feature engineering, and model evaluation techniques, demonstrating a holistic approach to data science.
- Recognized by Senior managers for dynamic thinking, effective problem-solving skills and able to handle high stress to complete the project deliverables and meet the deadlines.
- My background in Data Warehousing, Analytical tools, and Data Science equips me to maximize output and achieve business goals.

TECHNICAL SKILLS:

Languages: SQL, Python, PySpark and YAML.
Data Warehouse & RDBMS: Snowflake, MySQL.
Cloud Service: AWS (S3, Athena & Glue).
BI Tools: Qlik Sense, Microsoft Power BI & Tableau.
Modeling and Transformations: Data Build Tool.
Big Data: Hadoop, HDFS, MapReduce, Hive, Spark, Kafka, Apache Airflow.
Machine Learning: Linear Regression, Logistic Regression, KNN, Naïve Bayes, Decision Trees, Random Forest, Gradient Boosting, Adaptive Boosting, PCA, K-means clustering, Hierarchical Clustering, NLP & Time Series Forecasting.
Statistical Methods: Univariate, Bivariate and Multi variate Analysis, Hypothesis Testing.

WORK EXPERIENCE:

Associate Consultant, Capgemini
May 2021 -Present

- *Senior Data Engineer, Boehringer Ingelheim, Sept 2022-present (Snowflake, AWS, DBT Core, Airflow, GIT)*
  - Design and develop pipeline (stored procedures) to ingest the data from AWS S3 into Snowflake by modelling in DBT and scheduling the jobs in Airflow by traversing through all relevant schemas by incorporating SCD type 2 methodology to track historical changes.
  - Develop the views by joining the datasets based on the business logic and the nuanced requirements of Power BI reporting.
  - Perform unit and regression testing on the models and procedures and create testing documents for each initiative.

- *Snowflake Developer / Qlik Sense Analyst, Siemens Energy, Nov 2022-Sept 2023 (Snowflake, Qlik Sense, Tableau, GIT & AWS S3)*
  - Perform an Initial analysis of the Qlik Sense Applications to migrate it into Snowflake SQL and Tableau Dashboards.

- o Migrate the Qlik Data Load Script into Snowflake SQL with Exceptional Handling Snowflake by performing Data Wrangling.
- o Create Tasks and Stored Procedures for Daily Refresh of the Data.
- o Ingest flat files data from AWS S3 into snowflake by using Snowpipe, streams and stages.
- o Deploy the Code through GIT into Snowflake Production Environment.
- **Qlik Sense Developer, Intesa Sanpaolo, Oct 2021-Aug 2022 (Qlik Sense & Microsoft SQL Server)**
  - o Understand the Client Requirements to build the Applications.
  - o Create 3-tier architecture in Qlik Sense for ETL process with SQL Server as source and develop Qlik Sense Applications based on the Client Requirements.
  - o Performed Unit Testing and Regression Testing on the Applications.
    Achievements: Awarded with Rising star.

DATA SCIENCE PROJECTS:
- **Wine Sales Forecasting:**
  - o **Objective:** Create a model that can predict the sales of the Rose and Sparkling Wine by analysing its trend and seasonality and predict the sales of the future 12 months.
  - o **Tools & Techniques used:** Exploratory Data Analysis, Seasonal Decompose, ARIMA, SARIMA, Exponential Smoothing, Linear Regression, Moving Average, Naïve Forecast.
  - o **Outcome:** After creating different models on the sales data a Triple Exponential Smoothing Model with the least RMSE has been created that captures the trend and Seasonality of the data and is able to predict the future sales.
- **Election Exit Poll Prediction:**
  - o **Objective:** Create a model that can predict a voter's vote based on the given features that will be used for exit polls and predicting the overall wins and seats covered by a particular party.
  - o **Tools & Techniques used:** Exploratory Data Analysis, K-nearest neighbours, naïve bayes and Random Forest, Grid Search CV.
  - o **Outcome:** A random forest model is developed that can predict the voter's vote with an accuracy range of 80-87% with a recall and precision ranging from 70-80%.
- **Credit Card Retention Model:**
  - o **Objective:** Create a model that will help the bank to improve its services so that customers do not renounce their credit cards and give them the possible attributes that are causing the customer to renounce.
  - o **Tools & Techniques used:** Exploratory Data Analysis, SMOTE, Random Forest, Gradient Boosting and Ada Boosting.
  - o **Outcome:** An Adaptive Boosting model was developed that can identify the customers who are renouncing their credit cards and the possible features that are causing the renouncement's. The model has an accuracy range of 92-98% with a good precision and recall of 84-88%.
- **Text Analytics:**
  - o **Objective:** Analyse the speeches of the presidents and create a word cloud and identify the most used words in the speeches.
  - o **Tools & Techniques used:** nltk, text cleaning, word cloud.
  - o **Outcome:** The 3 speeches were analysed, cleaned and the most common words in the 3 speeches were identified and a word cloud of the 3 speeches were created.
- **Computer-active Prediction:**
  - o **Objective:** Build a linear equation to build a model to predict Portion of time (%) that CPU's runs in user mode and explore the effect of each feature on the time.
  - o **Tools & Techniques used:** Exploratory Data Analysis, Linear Regression.
  - o **Outcome:** A Linear Regression is model is created with a r-square of 62% and a linear equation is built that helps to identify the contribution of each attribute to the CPU's runs.
- **Women's contraceptive Project:**
  - o **Objective:** Build a model that can identify the women use of contraceptive methods studying the demographic and Socio-Economic factors.
  - o **Tools & Techniques used:** Exploratory Data Analysis, Logistic Regression.
  - o **Outcome:** A Logistic Regression model is developed with an accuracy of 64% and an AUC of 0.67.

EDUCATION:

Degree	Year	Percentage/CGPA	Institute
PGP- Data Science and Business Analytics	2023-pursuing	97.1%	Great Lakes & University of Texas
B.E Mechanical Engineering	2021	8.51	Sathyabama Institute of Science and Technology

PROFESSIONAL CERTIFICATIONS:

o Microsoft Certified: Power BI Data Analyst Associate (Certification Number: I432-2523).
o Microsoft Certified: Azure Data Fundamentals (Certification Number: I495-8030).

PROFESSIONAL COURSES:

o Mastering Big Data Analytics (Hadoop + Hive + Spark + Kafka), Great Learning.
o PySpark and AWS: Mastering Big Data with PySpark and AWS, Udemy.
o 2024 Mastering dbt (Data Build Tool)- From Beginner to Pro, Udemy.
o Apache Airflow | A Real-Time & Hands-On Course on Airflow (Airflow + Docker), Udemy.

Degree	Year	Percentage/CGPA	Institute
PGP- Data Science and Business Analytics	2023-pursuing	97.1%	Great Lakes & University of Texas
B.E Mechanical Engineering	2021	8.51	Sathyabama Institute of Science and Technology