Sagar Goswami | Data Science

goswami.sagar.d@gmail.com GoswamiSagarD goswami-sagar

*I am a Data Analyst looking for **Full-time positions** to build experience in **Big-Data Analytics / Machine Learning / AI / Statistical Optimization / Operations Research.** Having gained domain knowledge in multiple fields, I am looking for opportunities to put my skills to use in a professional environment.*

Education

M.S. Data Analytics Engineering	(3.89 of 4.00 GPA)	George Mason University	(2023)
B.E. Mechanical Engineering	(7.59 of 10 CGPA)	Gujarat Technological University	(2018)

Work Experience

Data Analytics and AI Analyst – Summer Intern

Sequent Inc., Somerset, NJ.

(February 2023 – Present), (June 2022 – August 2022)

- **Tested capabilities and limitations of Azure Synapse Analytics Platform** for automating entire Data Science projects. **Knowledge transfer to the analytics team** on usage and functionalities of Azure Synapse Platform, merits/demerits to using **Azure Synapse** vs. **Azure ML**. **(turn the page over for details.)**
- **Performed Data Mining on Patient Referrals Data** for a client project. Increased Prediction accuracy from 77% (previously) to 95% using various data encoding and pre-processing techniques. **(turn the page over for details.)**
- **Provided training to the staff** on **Python**, **Apache Spark**, **Azure Synapse**, Machine Learning, and Data Analytics Project Lifecycle.

Vice-President (Student Organization)

Horizon Tech, George Mason University, VA.

(November 2022 - Present)

- **Conduct and manage Data-Science Events, community projects**, **host a weekly tech podcast**, and **supervise and guide domain and operational teams** for the proper functioning of the student organization.

Technical Documentation Engineer

L&T - Defence and Aerospace, Powai-Mumbai.

(August 2019 – March 2020)

- Worked in a team to **develop Computer-Based Training Software** for K9 Vajra-T Self-Propelled-Howitzer.
- **Created a Hierarchical Data Model** for the Powertrain and Drive Train data with upwards of 20,000 parts, shortening the project timeline by more than four months.
- **Managed a team** of animators and created dashboards for tracking project progress on 3D Modelling, Animations, Graphics, UI/UX design, documentation, and query/issue tracking.

Programming / Tools

Python

pandas, NumPy, SciPy, Scikit-learn, PyTorch, TensorFlow, matplotlib, seaborn, Gurobi, OR-tools, beautifulsoup

R Language

tidyverse, ggplot, caret, leaflet

SQL/No-SQL Databases

MySQL, Oracle SQL, SQLite, MariaDB, MongoDB

Apache Spark

MLlib, SparkSQL, PySpark, SparkR, Spark on Databricks

Data Reporting and Business Intelligence

Tableau, Power BI

Key-Skills

Data Mining / Machine Learning / AI

Regression, Classification, Clustering, Time-series, Dimensionality Reduction, Deep Learning

Cloud Computing and Big-Data Analytics

Azure, AWS, Google Cloud, Databricks

Statistical Techniques

Optimization, Risk Analysis, NLP/GRG Algorithm, Network Models, Bayesian Statistics, Simulations, Stochastic Modeling

Data Visualization

Dashboards, Business Intelligence, Model Interpretation

Soft-Skills

Agile Development, Attention to detail, problem-solving, consistency, planning, communication

To learn more about my work, **turn the page to find my project portfolio**.

Project Portfolio (Sagar Goswami)

Projects from Professional Experience

Predicting Patient Admits for Hospice and Home-Health services based on referrals data obtained from Physicians/Healthcare facilities.

Sequent Inc., NJ.

- Analyzed Referrals Database to **predict patient admission for Home-Health and Hospice Services**.
- Tested various classification algorithms (Logistic Regression, Decision Trees, Random Forest, Naïve-Bayes, and Support Vector Machines) for predictions.
- Used **Python** and **sci-kit-learn** library on a remote machine on **Microsoft Azure**.
- **Developed a custom python library** (based on **pandas**, **sci-kit-learn**, **seaborn**) to **automate ML Model training, hyperparameter tuning, cross-validation and testing, computing, and comparing predictions accuracies/RMSE for various algorithms/models, making predictions and merging them to the dataset, generating plots for model performance (confusion matrix, error plots)**

Azure Synapse (Proof of Concept for Data Mining)

Sequent Inc., NJ.

- **Created a POC (Proof of Concept)** to test the capabilities and limitations of **Azure Synapse Platform** to carry out the entire Data Science Project, from data ingestion to data mining and reporting/dashboards.
- **Performed Data Exploration and Data Analysis** in **Synapse Spark** using **SparkSQL** and **PySpark Notebooks**. **Performed Data Mining on Census Data** (Correlation Tests, Trend Analysis and Future Projecting, Multi-Variate Regression, and Polynomial Regression) using **Spark ML**.
- **Integrated Azure Synapse with various data sources** (**Azure Data Lake**, External RDBMS, and **Microsoft SharePoint**) and **external reporting tools** (**PowerBI** and **Tableau**) to store data and prepare Dashboards/Reports to **visualize the insights obtained from Data Analytics and Mining**.
- **Automated the entire process** using **Azure Synapse Pipelines** and **Azure Data factory**.

Academic Projects

- **Image Classification using Convolutional Neural Networks** with **Python** and **Tensorflow** to differentiate images of cats and dogs
- **Analyzing and predicting traffic speeds** for New York City using Clustering, PCA, and Time-Series Analysis (**Apache Spark on Databricks**, NYC Department of Transportation API, **Spark MLlib**, ML Pipelines, Facebook's **FBProphet** library, **PySpark**, **SparkSQL**, **SparkR**)
- **Analyzing and predicting the Average Life Expectancy** of countries using WHO Open Data. (**R-Language**)
- **Designed and developed a relational database** for storing and maintaining the daily nutrition intake of users. The database stores various recipes, their nutritional contents, user information, their daily dietary targets/requirements, and meals consumed by users. (Relational Databases, **SQL**, **Oracle SQL Server**)
- **Predicting Bike-Sharing Service Demand** based on previous data history, and external factors like Weather, Holidays, etc. (**R-Language**, **Python**, Regression Models, Feature Selection Techniques, Tree Models, Ensemble Models, Deep Learning, **TensorFlow**, automation, Hypothesis-Testing)
- **Classifying Wines** based on their features (Data Mining, **Python**, **sci-kit-learn**)
- **Optimized the Distribution Network using Linear Programming and Heuristic Methods,** for an Automobile Manufacturing Company. (**Python**, **Gurobi**, **gurobipy API,** Linear Programming, Heuristic Optimization techniques)
- **Analyzing Means of Commute to Work and Public Transportation usage** in the United States using U.S. Census Bureau Data. (**Python**, **R-Language**, **SQL**, **OpenData API**)