SAGAR GOSWAMI

 @sagardgoswami  goswamisagard@gmail.com  https://**github**.com/**GoswamiSagarD**

OBJECTIVE

Looking for Internship opportunities (Jan-23) and full-time positions (May-23) to build experience in ML/AI projects. My long-term goal is to become a Data Scientist.

WORK EXPERIENCE

SEQUENT INC., SOMERSET-NJ. (SUMMER-2022)

Data Analytics and A.I. Analyst Intern

- **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.
- **Performed Data Mining on Patient Referrals Data for a client project**.
- **Provided training to the staff** on Python, Apache Spark, Azure Synapse, Machine Learning, and Data Analytics Project Lifecycle.

L&T – DEFENCE AND AEROSPACE (2019-2020)

Technical Documentation Engineer

- 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

PROJECTS

PREDICTING PATIENT ADMITS FOR HOSPICE AND HOME-HEALTH SERVICES BASED ON REFERRALS DATA OBTAINED FROM PHYSICIANS/HEALTHCARE FACILITIES.

Sequent Inc.

- 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 running 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.

- **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 Analytics in Synapse Spark** using **SparkSQL** and **PySpark Notebooks**. Performed **Data Mining on Census Data** (**Correlation Tests, Trend Analysis and Future Projecting, Multi-Variate Regression, 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 Dashboard/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

- Analyzing and predicting traffic speeds for New York City using Clustering, PCA, and Time-Series Analysis (**Apache Spark on Databricks**, NYC Department of Transportation Data, **Spark MLlib**, **ML Pipelines**, **Facebook's FBProphet library**, **PySpark**, **SparkSQL**, **RSpark**)
- Analyzing and predicting Average Life Expectancy of countries using WHO Open Data. (R-**Language**)
- Classifying Wines based on its features (**Data Mining**, **Python**, **sci-kit-learn**)
- Analyzing Means of Commute to Work, and Public Transportation usage in The United States using U.S. Census Bureau Data. (**Python**, R, **SQL**, **OpenData API**)

EDUCATION

GEORGE MASON UNIVERSITY

M.Sc. Data Analytics Engineering

GUJARAT TECHNOLOGICAL UNIVERSITY

B.E. Mechanical Engineering

KEY SKILLS

DATA MINING

DATA VISUALIZATION

MACHINE LEARNING / AI

(Regression, Classification, Clustering, Time-series Analysis, Feature Selection)

CLOUD COMPUTING, PARALLEL PROCESSING, AND BIG-DATA ANALYTICS

DATABASE DESIGN

STATISTICAL TECHNIQUES

(Optimization, Risk Analysis, NLP/GRG Algorithm, Network Models, Sequencing, Simulations, Stochastic Models)

PROGRAMMING LANGUAGES / TOOLS

PYTHON

(pandas, NumPy, SciPy, Scikit-learn, matplotlib, beautifulsoup, requests)

R LANGUAGE

(tidyverse, ggplot, caret, leaflet)

SQL

(MySQL, Oracle SQL)

APACHE SPARK

(MLlib, SparkSQL, PySpark, RSpark, Spark on Databricks)

NO-SQL DATABASES

(MongoDB)

TABLEAU, POWER BI, MS VISIO

MS OFFICE

(Access, Word, Excel, Powerpoint, Outlook)