SAGAR GOSWAMI

: goswamisagard@gmail.com : GoswamiSagarD : goswami-sagar

Looking for Internship opportunities (Jan-2023) and full-time positions (May-2023) to build experience in ML/AI/Big-Data projects. Currently focusing on Deep Learning / Statistical Optimizations / Predictive Analytics / NLP.

WORK EXPERIENCE

DATA ANALYTICS AND AI ANALYST – SUMMER INTERN

SEQUENT INC., SOMERSET NJ. (SUMMER-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.
- **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.

TECHNICAL DOCUMENTATION ENGINEER

L&T - DEFENCE AND AEROSPACE, POWAI-MUMBAI. (2019-2020)

- 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.

SECRETARY (DATA SCIENCE AND COMPUTATIONAL SOCIETY)

GEORGE MASON UNIVERSITY. (2022-2023)

EDUCATION

M.S. DATA ANALYTICS ENGINEERING (3.95/4.00) GEORGE MASON UNIVERSITY

(ML/AI/BIG-DATA/OPERATIONS RESEARCH/DEEP-LEARNING) (EXP. 2023)

B.E. MECHANICAL ENGINEERING (7.59/10) GUJARAT TECHNOLOGICAL UNIVERSITY

(MACHINE DESIGN/COMPUTER-AIDED-DESIGN/MECHANICS) (2014-2018)

PROJECTS

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 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., 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, MultiVariate 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 API, 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-Language, SQL, OpenData API)

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, Sequencing, Simulations, Stochastic Models)

DATA VISUALIZATION

RELATIONAL DATABASES

PROGRAMMING LANGUAGES / TOOLS

PYTHON

(pandas, NumPy, SciPy, Scikit-learn, pytorch, tensorflow, matplotlib, seaborn, gurobi, OR-tools, beautifulsoup)

R LANGUAGE

(tidyverse, ggplot, caret, leaflet)

SQL

(MySQL, Oracle SQL, SQLite, MariaDB)

APACHE SPARK

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

DATA REPORTING AND BUSINESS INTELLIGENCE

(Tableau, Power BI)

MS OFFICE

(Access, Word, Excel, Powerpoint, Outlook, Teams, Visio, Project)

NOTES