



PROFESSIONAL Summary

- **IBM Certificate Data Science & Machine Learning Professional** with **5+ years** of experience specializing in **Data Science, Nanofabrication, Nanoelectronics**, Medical Image Analysis, and Telecom Data Analytics using **Python Machine Learning & Deep Learning** Techniques
- Extensively involved in **Data preparation on large dataset with distributed database, Exploratory analysis**, Feature engineering using Supervised and unsupervised modeling using **Optimization algorithm** to get accurate model.
- Good understanding of **Agile** and **Waterfall** Software Development Life Cycle (**SDLC**) Methodologies.
- Working knowledge of various Python libraries such as **NumPy, SciPy** for mathematical calculations, **Pandas** for data preprocessing/wrangling, and **Matplotlib** for **data visualization**.
- Capable in writing, editing and debugging code using **SAS, SAS/Base** and **SAS/Macro**.
- Performed multiple **Data Mining** techniques and derive new insight from the data.
- Performed preliminary **data analysis** using descriptive statistics and handled anomalies such as removing duplicates and imputing missing values.

PROFESSIONAL EXPERIENCE

Dallas Data Science Academy – *Remote Data Scientist Intern*; Dallas, TX

September 2021 – Present

- Designed and developed state-of-art deep-learning/machine-learning algorithms for analyzing the image and video data among others. Experience with **Tensor Flow, Theano, Keras**, and other **Deep learning Frameworks**, optimization algorithm.
- Extensively used open-source tools - **R Studio(R)** and **Spyder (Python)** for statistical analysis and building **machine learning** algorithms. Extensively used **Azure Machine Learning** to set up the experiments and create Web services for predictive analytics
- Implemented predictive models using machine learning algorithms linear regression, boosting algorithms and performed in-depth analysis, applied concepts of **R-squared, RMSE, P-value**, in the evaluation stage to extract interesting findings through comparisons.
- Designed easy-to-follow **visualizations** using **Tableau** and published dashboards, stories on web and desktop platforms.
- Performed feature scaling, feature engineering, and statistical modeling.
- Built Artificial Neural Network using **Tensor Flow** in **Python** to identify the customer's probability of canceling the connections.
- Created multiple Logistic Regression models that helped fine-tune fabrication process cycle times reduced from 6 hrs to 30 minutes

University of Missouri – Kansas City – *Data Scientist/GRA/GTA*; Kansas City, MO

August 2016 – Present

- Implemented and followed **Agile** development methodology within the cross-functional team and acted as a liaison between the business user group and the technical team.
- Extracted, transformed and loaded **ETL** datasets with basic modeling variables for each segment type using Proc Import and **SQL** queries, imported data from **Oracle** database into **SAS** files.
- Involved in various packages in **R** and **python-like ggplot, Pandas, NumPy, SciPy** and **Matplotlib**.
- Extract, transform, and load data from the source system to **Azure** data storage using **Azure** data factory, **Spark SQL**.
- Visualizing and presenting dashboards to stakeholders using **Tableau & ggplot** by utilizing various plotting techniques.
- Good Working Knowledge with **SourceTree** with **Git**.
- Pioneered the use of **A/B testing** to recommend strategies.
- Surface analysis of materials by Scanning Electron Microscope (SEM), Atomic Force Microscope (AFM), device fabrication with Electron beam lithography (EBL), wet and dry etching, mask pattern, metal evaporation, worked in the cleanroom

Bright Mind Solutions – *Data Scientist*; Grand Forks, ND

August 2015 – August 2016

- Queried, cleaned, and analyzed historical sales data of 200+ products; utilized **time-series forecast** models to forecast sales; automated data collection and ensured **ETL** code was in adherence to business requirements
- Responsible for gathering report requirements, collecting data using SQL Script and creating **Power BI** Reports. Used **Power BI** Power Pivot to develop a **data analysis** prototype and used **Power View** and **Power Map** to visualize reports.
- Worked on **Jenkins** to implement continuous integration (**CI**) and Continuous deployment (**CD**) processes.
- Wrote stored procedures and User Define Scalar Functions to be used in the **SSIS** packages and **SQL** scripts.

EDUCATION

University of Missouri – Kansas City, MO

August 2016 – December 2022

Ph.D. in Electrical Engineering
Major in Data Science, Nanofabrication
GPA: 3.84/4.00

University of North Dakota, Grand Forks, ND

Master's in Electrical Engineering

Major in Data Science

GPA: 4.00/4.00

January 2012 – August 2015

Islamic University of Technology, Dhaka, Bangladesh

Bachelor of Electrical & Electronics Engineering

Major in Data Analytics

GPA: 3.67/4.00

December 2004 – November 2008

SKILLS

- Python, Tensorflow, Keras
- Tableau, Power BI
- Pandas, Numpy, Sklearn, Pytorch, PyCharm, Matplotlib, SciPy Ggplot
- SQL Server, PostgreSQL, SQL
- Classification (Naïve Bayes, KNN, SVM, Decision Tree, Random Forest, etc.)
- Regression (Linear, Multiple Linear, Ridge, Lasso, Random Forest, SVR, etc.)
- Clustering (K-Means, Hierarchical)
- Machine Learning, Deep Learning (RNN, CNN, Transfer Learning)
- Data Mining, Data Visualization, Data optimization, Tableau
- Statistical Analysis, R
- Amazon Web Service (AWS), Azure
- Optimization Algorithm (Adam, Gradient Descent)
- SDLC/Agile/Waterfall
- Time series Forecasting
- A/B testing, ETL
- Natural Language Processing (NLP)

Links

GitHub: <https://github.com/sehtab>
Azure ML Studio: <https://gallery.azure.ai/browse?s=Sehtab%20Hossain>
Tableau: <https://public.tableau.com/app/profile/sehtab.hossain>
LinkedIn: www.linkedin.com/in/sehtab-hossain-15a04892

Selected Projects

- Breast Cancer Classification: <https://github.com/sehtab/Breast-Cancer-Classification>
- Email Spam Classification: <https://github.com/sehtab/Email-Spam-Classification>
- Credit Card Fraud Detection: <https://github.com/sehtab/Credit-Card-Fraud-Detection>
- Stock Predictions: <https://github.com/sehtab/Stock-Prediction>
- Recommender System: <https://github.com/sehtab/Recommender-System>
- Sentiment Analysis: <https://github.com/sehtab/Sentiment-Analysis>
- Face Detection: <https://github.com/sehtab/Computer-Vision/tree/main/Face%20Detection>
- Object Detection Using YOLO: <https://github.com/sehtab/Computer-Vision/tree/main/yolo>
- Real-Time Object Detection and Enumeration of object: <https://github.com/sehtab/Computer-Vision/blob/main/liveobjectdetection.py>

Publications

Google Scholar Link: <https://scholar.google.com/citations?user=V17Dws0AAAAJ&hl=en&oi=ao>