SUMAIA P SHUPTI Vancouver, BC

sumaia.p.shupti@gmail.com ◆ (672) 999-7037 ◆ sumaiaparveen.github.io/portfolio

GitHub: github.com/SumaiaParveen
LinkedIn: linkedin.com/in/sumaia-p/

Key Skills

Data Collection, Data Wrangling, Feature Engineering, Exploratory Data Analysis, Data Visualization, Predictive Analytics, Statistical Analysis, Machine Learning, Regression, Classification, Clustering, NLP, Web scraping, ETL, RDBMS, PostgreSQL, Google BigQuery, Sage ERP Systems

Languages

Python
SQL
C, MATLAB

Libraries

Pandas, NumPy, BeautifulSoup, Missingno, Datetime, Matplotlib, Seaborn, Bokeh, Plotly, Folium, ipython-sql, Scikit-learn, SciPy, GeoPy, Yellowbrick, Streamlit, Pyproj, Bqplot, Altair, Nltk, textblob, spaCy, AutoKeras, LazyPredict, PyCaret

Reporting Tools

Tableau
MS Excel
Power BI

Other

AWS Sagemaker, SAS

Recent Projects

Web App: Natural Language Processing with spaCy Link, App

- The app takes input from user, pre-processes text, shows tokens and lemmas, shows common words of two documents and cosine similarity, provides extractive summarization using *Gensim*, *Sumy*.
- *Pandas, Nltk, Spacy, spacy_streamlit, Genism, Sumy, Sklearn, Wordcloud, Streamlit, Heroku*

ML Project: Binary-Classification of Health Conditions Link

- Employed *Lazypredict, Xgboost, Lightgbm, Catboost* and *EasyEnsemble* models to classify the labels, examined evaluation metrics *ROC_AUC, Precision-Recall AUC, Precision, Recall, F1 and Accuracy Score*, and selected a suitable model for the problem.
- *Pandas, Numpy, Sklearn, SciPy, Yellowbrick, Mlxtend*

Web App: Prediction of Air Quality Index of Los Angeles Link, App

- Scraped data using *BeautifulSoup* from multiple sources and combined in *Pandas*, optimized *Random Forest Regressor* using RandomizedsearchCV to reach the best model, deployed on Heroku.
- *Requests, BeautifulSoup, Sklearn, SciPy, Lazypreict, Lightgbm, Yellowbrick, Streamlit, Heroku*

ML Project: k-means Clustering of Dhaka Districts Link

- Scraped data using BeautifulSoup, utilized Foursquare API for names and coordinates of venues, employed k-means clustering to segment Dhaka division, analyzed results and spotted a suitable area to open a restaurant in the city.
- *Pandas, Numpy, Sklearn, Folium, Geopy, Urllib, BeautifulSoup, Foursquare*

Web App: Prediction of Human Life Expectancy Link, App

- Employed *ExtraTrees* Regressor using *RandomizedsearchCV* to reach the best model that takes input from the user and predict life span in years of people from 193 countries.

Please see more projects-- **sumaiaparveen.github.io/portfolio/projects/**

Education

Master of Science: Mathematics — 2017
University of British Columbia, Vancouver BC
Project: Squeeze Flow of Viscoplastic Fluids: A Variational Approach

Bachelor of Science: Electrical & Electronic Engineering — 2013
North South University, Dhaka, Bangladesh
Thesis: Finite Volume Method Simulation of Blood flow in a Model Arterial Aneurysm

Experiences

Machine Learning Intern--- Tech For Good Inc, Boston, MA, US — 03/2021-Present
Data Analysis
- Collected virtual machine pricing data using Google Cloud Billing API and Azure Retail Price API.
- Wrote scripts to pull hourly and daily data and integrate with previous timestamps' datasets.
- Extracted useful data from JSON and converted it to tabular format using Pandas.
- Analyzed data on Google BigQuery, pulled data to Pandas from BigQuery using API.
- Visualized BigQuery data on Google Data Studio and JupyterLab.

Predictive Analytics
- Collected data from various sources using AWS CLI, AWS API, and several Python packages i.e., boto3.
- Wrangled, preprocessed, analyzed, post-processed, and visualized the data on JulyterLab and Tableau.
- Performed regular and time-series feature engineering and hyperparameter tuning.
- Built various regression, time-series, and clustering models including SARIMAX, Prophet, bagging/boosting algorithms, autoML i.e., PyCaret and LazyPredict, some deep learning models i.e., LSTM, BiLSTM, GRU, AutoKeras, etc. to predict the desired outcome.

Supply Chain Data Analyst--- Ion Cure Tech (Remote) — 01/2021-03/2021-
- Communicated with stakeholders and designed questionnaires as per the needs of the clients
- Collected data using Python web-scraping and google forms
- Performed data wrangling in Python and designed database schema for dashboard developments in Power BI
- Enabled leadership to identify trends and generate actionable insights more efficiently by developing interactive data visualizations

Data Science Advanced Bootcamp, DPhi — 06/2020-08/2020
- Data preparation and Feature Engineering, End-to-End ML Model Building pipeline
- Time Series Analysis, Natural Language Processing

Graduate Teaching Assistant—University of British Columbia, Vancouver, BC, Canada — 09/2015-08/2017
- Taught a class of 35 students, prepared lecture notes, conducted tutoring one-to-one and in groups, midterms, and exams, invigilated tests.
- Courses assisted: Integral Calculus with Applications to Physical Sciences and Engineering, Multivariable Calculus, Calculus (Engineering Science), and Applied Linear Algebra.

Laboratory Instructor—North South University, Dhaka, Bangladesh — 09/2014-04/2015, 05/2013-08/2013
- Prepared Laboratory Manuals, delivered lectures, demonstrated experiments, evaluated students' lab performance and lab reports, conducted and graded lab exams.
- Taught Design and Simulation of VLSI Layouts and solving Engineering Mathematics problems using Wolfram Mathematica to classes of 3rd and 4th year undergraduate students
- Courses taught: Assembly programming in Microprocessor and Interfacing Lab, Introduction to VLSI Design Lab, Engineering Mathematics Lab

Research Assistant—North South University, Dhaka, Bangladesh 09/2013-08/2014

- Participated in designing study (problem formulation, computation), simulation, acquisition of data, critical analysis, and result interpretation
- Presented data and research findings in conferences
- Wrote research articles in international peer-reviewed journals as well as technical replies to the questions raised by reviewers
- Co-supervised a thesis group; helped with code modification, mathematical computation, and writing articles
- Assisted supervisor to review journal articles
- Worked both independently and in a team

Teaching Assistant—North South University, Dhaka, Bangladesh 09/2011-04/2013

- Conducted tutorial sessions, graded homework, programming assignments, and exams
- Courses assisted: Introduction to VLSI Design, Calculus and Analytical Geometry II, Electrical Services and Engineering Economics and Management

Certifications & Independent Studies

Data Science Advanced Bootcamp, DPhi 08/2020
200+ hours coursework, 3 capstone projects over more than 7 weeks on end-to-end ML/AI model building and deployment.

Applied Machine Learning in Python, University of Michigan 02/2020
Skills: Regularization, feature scaling, and cross-validation to avoid under- and overfitting, evaluation metrics and interpretation of results, model optimization, model selection, avoiding instances of data leakage.

Data Wrangling, Analysis and AB Testing with SQL, University of California, Davis 01/2020
Skills: Coalesce nulls, identifying unreliable data, defining new metrics tied to a business value, mapping out joins, aggregating a proportion metric and run the results through an AB testing calculator tool.
Project: A/B testing of an e-Commerce site.

Analyze Data with SQL, Codecademy 12/2019
Skills: Joining multiple tables, window functions to compute across a set of table rows.
Projects: Finding the best dinner spot in NY, identifying potentially risky transactions, customer segmentation, finding a funny tagline from the previous burger orders, data exploration of the Metropolitan Museum of Art and Hacker News, analyze the ledger data, calculating churn rates, analyze marketing funnels, marketing attribution: mapping customer journey

Analyze Business Metrics with SQL, Codecademy 12/2019
Skills: Advanced aggregation and exploration of KPIs.
Projects: Summarize a meal delivery app data. Producing KPIs of a video game app.

Learn SQL, Codecademy 12/2019
Skills: Manipulating data and building queries to communicate multiple tables.
Projects: Analyzing trends in a startup, analyzing ride-sharing app data.

Data Science Professional Certificate, IBM 11/2019
Tools: Jupyter/JupyterLab, GitHub, R Studio, and Watson Studio
Skills: Accessing databases using Python, handling missing values, binning, grouping, EDA, data standardization, descriptive statistics, ANOVA, model evaluation, Linear, grid-search, model selection, prediction, model refinement, handling over-fitting, under-fitting.

Projects: Random album generator (recommendation system), predict housing prices (regression), best classifier model(classification), battle of neighborhoods (clustering).

Analyzing and Visualizing Data with Excel, Microsoft, edX
06/2018

Skills: VLOOKUP, data modelling, mash-up of data from multiple sources, measures, advanced text query

Analyzing and Visualizing Data with Power BI, Microsoft, edX
06/2018

Skills: Transforming data (column operations, query groups and parameters), data modelling in Power BI desktop (calculated columns, measures, time intelligence, include/exclude, grouping/binning), Data Visualization in Power BI Desktop (Analytics Pane, Clustering, Map Visualizations, Default Summarization & Categorization, Custom Hierarchies)

Introduction to Data Analysis using Excel, Microsoft, edX
02/2017

Skills: Data modelling, multi-table pivots, profitability analysis, year to year analysis, reporting hierarchical data.

Publications

- Pulsatile Non-Newtonian Fluid Flows in a Model Aneurysm with Oscillating Wall. Frontiers in Mechanical Engineering, 2017.
- Rheological Behavior of Physiological Pulsatile Flow Through a Model Arterial Stenosis with Moving Wall. Journal of Fluids,2015.
- Pulsatile Non-Newtonian Laminar Blood Flows Through Arterial Double Stenoses. Journal of Fluids, 2014.
- Laminar Blood Flow Through a Model of Arterial Stenosis with Oscillating Wall. International Journal of Fluid Mechanics Research, 2014.
- Pulsatile Laminar Flows in A Dilated Channel Using Cartesian Curvilinear Coordinates. Universal Journal of Mechanical Engineering, 2013.
- Physiological Pulsatile Laminar Blood Flow in a Compliant Arterial Aneurysm. 18th International Mathematics Conference- 2013.