Vrajesh Patel

1013-35 Fountainhead Rd, North York, ON, Canada

[LinkedIn](https://www.linkedin.com/in/vrajesh-patel-8241b41a3/) [Portfolio](https://vrajeshpatel.herokuapp.com/) [GitHub](https://github.com/VrajeshPatel20) Email: [vrajesh1@my.yorku.ca](mailto:vrajesh1@my.yorku.ca) Phone: +1 (647) 609-8945

HIGHLIGHT OF QUALIFICATIONS

* Results-driven **data scientist** with profound knowledge of mathematics, statistics, and machine learning concepts with a knack for finance.
* Hands-on experience building a modern ETL Data Pipeline using Airflow and Azure DataBricks.
* Detail-oriented self-starter with excellent communication skills, problem-solving skills and thrive in a fast-paced environment.

TECHNICAL SKILLS

**Programming Languages**: Python, Java, SQL, HTML, CSS, JavaScript

**Analytics Libraries**: PySpark, Pandas, NumPy, Matplotlib, Seaborn, Plotly, SAP BI Power Tool

**Machine Learning Toolkit**: PyTorch, Scikit-learn, Keras - TensorFlow, AWS

**Software & Tools**: Azure Databricks, Airflow, Matlab, Tableau, Git, Docker, PyCharm, VS Code, Colab, Jupyter

WORK EXPERIENCE

**Cloud Data Engineer Intern**

RBC | Jan 2023 – Present

* Configured Apache Airflow as the workflow manager leveraging Azure Databricks as the executor for Airflow tasks, harnessing sparks distributed processing capabilities to improve performance over large-scale datasets. While designing and defining tasks for Airflow Dags to perform data movement and distribution.
* Implemented the Medallion structure from Databricks to enable isolation and resource management for multiple end users. While utilizing Airflows Dags to have better data lineage and failure recovery protocols.
* Leveraged modern databricks functionalities like upserts and autoloader to enhance performance while processing large scale datasets while performing ETL.

**Enterprise Architect and Business Innovation Intern**

Toronto Hydro | Sept 2021 – May 2022

* Financial Analysis of total expenditure of organizations' small intakes by generating KPI reports that assess cumulative expenses, execution timelines, number of intakes, overspending on work orders, and distribution of initiatives as per groups, threshold levels, and more using a combination of Alteryx, Tableau, and Excel VBA.
* Build KPI Report that keeps track of the overall financial expenditure over all the departments as well as categorized the actual amounts in terms of the amount approved, in development, and the buffer to the budget.

RELEVANT PROJECTS

**Market Analysis (Social Influence on Stocks and Cryptos) [Individual Project, Programming Languages: Python, HTML, CSS, JavaScript] {**[**Website**](https://social-influence-analysis.herokuapp.com/)**,** [**GitHub**](https://github.com/VrajeshPatel20/Market-Analysis)**}**

* Started by mining data using either API or requests library from python to get real-time data. Retrieved over 3000 posts (Reddit) and tweets (Twitter) along with current stock and crypto prices from various sources as dataframe.
* Analyzed the retrieved posts and tweets by cross-referencing the names of stocks and cryptos with the latest posts (Reddit) and tweets (Twitter). Trained a model for sentiment analysis using a simple linear and embedded layer from the ‘nn’ Library from PyTorch and achieved 80% accuracy.
* Visualized results obtained from the data of the top trending stocks.
* Performed linear regression on the results to get a relation between the influence of positive posts and the influence of negative posts to the changes experienced in the stock or crypto with 50% accuracy due to insufficient data.
* Presented the results on a website created on HTML, CSS, and JavaScript.

**Face Mask Detection [Individual Project, Programming Languages: Python] {**[**GitHub**](https://github.com/VrajeshPatel20/FaceMask-Detection)**}**

* Obtained the updated dataset from Kaggle with over 12000 different images and processed it using PyTorch.
* Trained a model with multivariate linear regression with over 97% validation accuracy and 9.8% validation loss.

**Heart Attack Prediction [Individual Project, Programming Languages: Python, Tableau] {**[**GitHub**](https://github.com/VrajeshPatel20/HeartAttack-Predictions)**}**

* Retrieved data from Kaggle and performed data analysis. Used Linear Regression using Deep Learning concepts.
* Generated plots in python and created an analysis dashboard on Tableau.

EDUCATION

**York University**, Bachelor of Science, Computer Science Jan 2019- Present