tallal.usman@gmail.com

Tallal Usman

857-472-1408

github.com/tallalUsman

tallal-uoft.github.io

linkedin.com/in/t-usman/

EXPERIENCE

Extend *San Francisco, CA, U.S.A.*

Senior Data Scientist **Dec 2021 – Present**

- Spearhead team working on ML, engineering and product initiatives at a growth-stage 300-person Series C Fintech startup.
- Partner with Finance, Risk, CX and other stakeholders to automate and enhance analysis using Tableau and Python.
  - Crafted and automated financial forecasts for 500+ merchants using time series models such as ARIMA and Prophet.
  - Partner with CRO in designing Tableau dashboards to track product insights relating to churn, retention, conversion, etc..
- Utilized A/B testing and regression analysis to enhance user conversion and underscore our product's value to merchants.
  - Champion A/B testing to elevate conversion and spearhead development of the infrastructure for experimentation.
- Built machine learning predictive models to improve revenue retention and reduce costs using scikit-learn and Snowpark.
  - Deployed a logistic regression model in AWS Sagemaker utilizing sales and alt. data, reducing churn by >$25M in 2022.
  - Developed a Random forest estimator to predict claim volumes and schedule personnel accordingly, reducing cost by 15%.
  - Utilize open-source computer vision for claims, streamlining approval, and fraud detection, reducing manual costs.
- Construct scalable data and MLOps pipelines to automate, deploy and share predictive analytic using Pyspark and AWS.
  - Act as team analytics engineer to set up data pipelines and reverse-ETL to Salesforce using AWS, DBT, Snowflake.
  - Invent our Python library, extend_analytics to streamline collaboration and deployment with CI/CD, AWS CodeArtifact.

Charles River Associates *Boston, MA, U.S.A.*

Associate **Sept 2019 – Dec 2021**

- Led teams as a data and analytics consultant on competition to Fortune 500 using transaction, financial, etc data.
- Presented memos, decks, and visualizations of findings to technical/non-technical stakeholders such as C-suite executives.
- Reduced client cost by $150K through automating geospatial mapping analyses of client traffic data using geopandas and GIS.
- Employed hypothesis testing, experimentation, MMM, and panel data regressions to provide competitive insights.
- Forecasted growth of a two-sided social media platform by using discrete choice modelling and machine learning.
- Removed bottlenecks by speeding up big data analysis using vectorization and multi-processing through Ray and joblib.

University of Toronto *Toronto, ON, Canada*

Research Assistant, Department of Economics **May 2018 – Aug 2019**

- Led RA's in analyzing auction data for anomalies using isolation forests and NLP on text documents using spaCy.

EDUCATION

University of Toronto *Toronto, ON, Canada*

Hons. Bachelors in Economics and Statistics **Jun 2019**

Arthur Hosios Scholarship in Economics: Awarded annually for the best original research thesis in empirical economics.

DBT Certified Analytics Engineer **March 2023**

PROJECTS

The Office Reboot: LLM cut

- Generate novel dialogue for the Office by fine-tuning a LLM on the scripts from the office on a T4 GPU.
- Used QLora and Huggingface to speed up fine-tuning and store/serve the model for inference.

Deploying Deep Learning applications in Sports betting

- Designed a streaming MLOps pipeline on Google Cloud (GCP) using CI/CD, Docker, Artifact Registry, Databricks, Airflow, DBT written in SQL, Python, Pyspark, Go, for ELT, feature engineering and predicting football outcomes.
- Employed a deep learning Pytorch model as well as Multinomial-logit, Random forests, XGBoost etc. to assess accuracy.

Leveraging NLP on alternate data for trading

- Predict bearish/bullish trends in stock prices by tracking sentiments for companies on the S&P500 using a deep learning Fin-BERT transformer leveraging scraped Google News and Reddit *wallstreebets* data.

TECHNOLOGIES & SKILLS

Programming: Python, SQL, Spark, Go, jinja, R, STATA, MATLAB

Technologies: DBT, Mlflow, AWS/GCP, Snowflake, Databricks, Ray, Airflow, Docker, Tableau, Kubernetes, Terraform

Packages: pandas, numpy, sci-kit learn, seaborn, polars, Pytorch, Duckdb, tidyverse, CARAT, E1071, LightGBM, MLlib, SHAP