

Yashwant Singh

San Francisco, CA | Phone: +1 4708151406 | ysingh64@gatech.edu

EDUCATION

Georgia Institute of Technology

Master of Science in Analytics

Atlanta, GA

Aug 2021-Dec 2022

Indian Institute of Technology, Kanpur

B.Tech - M.Tech dual in Materials Science and Engineering

Kanpur, India

Jul 2013 - Jun 2018

WORK EXPERIENCE

EQUIFAX | Data Science Intern

May 2022 – Aug 2022 | Atlanta, GA

- Developed an automated RCA process using python to identify top features impacting credit scores on Google Cloud Platform to preserve accuracy of data/credit scores for over 300 million people
- Implemented logistic regression and XGBoost models to identify important attributes for over 600 credit models, which contribute to \$3.5Billion of Equifax's revenue

SHARECHAT | Data Scientist - Product Analytics

Sep 2019 – Jun 2021 | Bangalore, India

- Managed a team of 5 to migrate CleverTap (analytics and notifications tool) to Looker
- Deployed Looker models using Advanced SQL from scratch while retaining all CleverTap functionalities, thereby reducing cost by 90%, translating to savings worth \$27000/month
- Collaborated with ML Engineers to build an in-house Not Suitable For Work (NSFW) model to replace a 3rd party model; saved \$25000/month
- Designed and analysed over 20 A/B tests to derive actionable insights to improve the retention by 3% for the app with over 200 million monthly active users

MYNTRA | Data Analyst

Jul 2018 – Aug 2019 | Bangalore, India

- Conceptualised a tool using python and SQL to improve line balancing and produce alerts to reduce leakage in customer order consolidation, resulting in a 2% decrease in leakage and annual savings worth \$250,000
- Facilitated warehouse teams with implementing changes in the order-fulfilment process, resulting in shrinking order-to-pack time from 4.0 hours to 3.0 hours

PROJECTS

LEAD PRIORITIZATION | Pilot Flying J - Applied Analytics Practicum

Aug 2022 – Dec 2022

- Implemented logistic regression/XGBoost models to predict propensity of lead filling fuel using 300 features for 470k leads
- Engineered 800 additional features (mean, mode, skew, kurtosis) using tsfresh, executed SHAP Analysis to determine feature importance and assessed model performance using recall

REDDIT - TOPICS OVER TIME | Link - [Course Project](#)

Aug 2022 – Dec 2022

- Employed Reddit Pushshift API to collect 300k posts/2.25 M comments for 3 subreddits, trained LDA/BERTopic models
- Deployed a web app to visualise how topics and sentiments change over time

FANTASY PREMIER LEAGUE POINT PREDICTION | Link - [Hobby Project](#)

Jun 2022 – Aug 2022

- Developed an ML-based app on streamlit allowing users to select players and predict points using linear regression and XGBoost models, compare players on metrics including xG, goals, assists
- Scraped player-level data using Beautiful Soup; Ranked in the top 1 percentile among 10M+ players in the world

PERSONALISED ITEMS RECOMMENDATION | Link - [Course Project](#)

Jan 2022 – Apr 2022

- Spearheaded a team of 4 to implement recommendation algorithms (CTDNE, JODIE) based on 7M reviews on Amazon

SKILLS

Programming: Python (Pandas, Numpy, SciPy, Scikit-learn, PySpark, Tensorflow, Matplotlib, Streamlit, Stellargraph, UMAP, HDBSCAN, tsfresh, SHAP, BERTopic), SQL, Git, Github, R, Spark, Hadoop, Looker, Tableau, D3

Statistical/Analytical Methods: A/B Testing, Hypothesis Testing, Time Series Modelling, Predictive Modelling, Data Visualization, Statistical Modelling, Clustering, Sentiment Analysis, Web Scraping, Exploratory Data Analysis