

Yashwant Singh

Chicago, IL | Phone: +1 4708151406 | yashwantsingh6394@gmail.com | [linkedin.com/in/ysingh63/](https://www.linkedin.com/in/ysingh63/)

WORK EXPERIENCE

DECISION SPOT | Data Scientist

Mar 2023 – present | Chicago, IL

- Developed an advanced demand forecasting model using Temporal Fusion Transformer for a Fortune 500 client, improving forecast accuracy, and resulting in a \$10M reduction in manufacturing costs by accurately predicting demand.
- Automated the data pipeline on Databricks, handling over 100GB of raw data
- Designed and implemented a RAG-based Q&A system using SBERT embeddings and LLM-GPT4
- Leveraged historical Q&A data/App documentation for faster response generation, improving operational efficiency by 56%
- Built an ARIMA-based anomaly detection model to identify significant \$/BBL fluctuations in the client's supply chain, detecting \$25M in financial anomalies.

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

Sep 2019 – Jun 2021 | Bangalore, India

- Designed and implemented two predictive models to identify and reduce NSFW content, leveraging CNNs for image embeddings and fine-tuned BERT models for text, leading to a 95% reduction in NSFW content views
- Built an adaptive content filtering pipeline that incorporated active learning from manual reviews, improving model accuracy and reducing the prevalence of low-quality content by discarding or flagging NSFW posts
- Managed a team of 5 to deploy Looker models using Advanced SQL from scratch while retaining CleverTap functionalities, thereby reducing cost by 90%, translating to savings worth \$27000/month

MYNTRA | Data Scientist

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-fulfillment 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 using tsfresh, executed SHAP Analysis to determine feature importance

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 visualize how topics and sentiments change over time

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

EDUCATION

Georgia Institute of Technology, Master of Science in Analytics

Aug 2021-Dec 2022

Indian Institute of Technology, Kanpur, B.Tech - M.Tech in Materials Science and Engineering

Jul 2013 - Jun 2018

SKILLS

Programming: Python (Pandas, Numpy, SciPy, Scikit-learn, cuDF, PySpark, Pytorch, Tensorflow, Matplotlib, LangChain, Stellargraph, FAISS, SBERT, tsfresh, SHAP, BERTopic), Snowflake, SQL, Git, Github, R, Spark, Looker, Tableau, PowerBI

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