

Vedant Ghavate

+1 (747) 305 0428 | vedantghavate259@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

WORK EXPERIENCE

Business Intelligence Developer, *Tata Consultancy Services*

Feb 2021 - Aug 2022

- Provided Cross-functional support for Data Warehousing, ETL, and Migration for three projects (80/20 Analysis, SPS Pricing, and Intercompany) as an end-to-end Snowflake and Tableau resource from development to maintenance.
- Developed over 50 SQL views to categorize accounts in Snowflake on six levels for AR/AP/GRIR on six SAP systems.
- Designed SQL scripts for data loads and scheduled 12 IICS jobs to automate migration and reduce 95% manual intervention.
- Identified 82% of dormant accounts through six Tableau Dashboards, resulting in business-level recategorizing of accounts.
- Prepared 60+ technical documents detailing maintenance and debugging scope for the migration pipeline.
- Investigated the root cause of inflated sales and identified vendors to be excluded to remove abnormal data from Dashboards.
- Achieved a turnaround time from 2 weeks to 36 hours to resolve data related issues with Tableau Dashboard.

Data Analyst, *Yash Dairy & Milk Products*

May 2020 - Dec 2020

- Built a Python-based parser for existing spreadsheets and a report-generation system for a dairy business with 650 accounts.
- Saved 97% of the time required by replacing manual entry and allowing analysis of daily deliveries
- Engineered Maps, Line and Bubble Charts for reviewing monthly data, predicting order volume, and seasonal sales.
- Analytics overhaul resulted in identifying reducing 90% of milk wastage and identifying 45% of new customers.

TECHNICAL SKILLS

Programming Languages: Python, R, C, C++

Tools: Tableau, Airflow, Spark, PowerBI, ArcGIS, Kafka, Excel, Looker, Hive, Klipfolio

DBMS: MySQL, MongoDB, Google Big Query, Snowflake

Cloud Technologies: AWS (EC2 Instance, S3 Bucket)

Frameworks: Numpy, Scipy, Scikit-Learn, Tensorflow, Keras, Pytorch, Pandas, Nltk, Matplotlib, Folium, Statsmodels, Spacy, Scrapy, Lightgbm, Theano, Ramp, Pipenv, Caffe2, Chainer, langchain

EDUCATION

Master of Science - Big Data Analytics

San Diego State University

Aug 2022 - May 2024

Bachelor of Engineering - Information Technology *Savitribai Phule Pune University*

Aug 2016 - May 2020

Coursework: Predictive Models, Data Science, Data Visualization, Statistical Methods, Reinforcement Learning, Business Analytics, Artificial Intelligence, Big Data Tools, Data Mining, Neural Networks, Computer Vision, Google Earth Engine, Prompt Engineering, Diffusion Models, Transformers

RESEARCH EXPERIENCE

Graduate Assistant, *San Diego State University Research Foundation*

Feb 2023 - May 2024

- **War in Ukraine and Energy Stock Market** - Collected ~15000 tweets through Tweepy, developed a data pipeline for preprocessing and sentiment analysis, and predicted stock price performance using Natural Language Processing (NLP); The energy stock market data was trained through a time series transformer with Twitter sentiment as an exogenous factor.
- **Generative Question and Answer System** - Developing a Large Language Model (LLM) and BERT-based Model to summarize and extract insights based on a dataset of over 1600 published research papers.
- Used unsupervised topic modeling algorithm 'lib2vec' to categorize research papers and extract the paper's central idea.
- System Engineer for The Language Acquisition Resource Center - Streamlined Selenium and Python scripts to automate audio prompts' validation and data entry, overachieving the initial migration target of 3 languages to 16 languages.
- Newsletter Editor for 'AI4Business Lab' to procure funding for the lab, organized presentations for guest speakers, and oversaw research.

Disparity in Affordable Healthcare Insurance Across America

Jan 2024 - Present

- Assembled Map based Dashboards with Pie charts, to highlight plan prices based on population and income and to identify counties offering plans based on members.
- Performed EDA of ~14 GB datasets across 2014-2024, to extract significant corrections and engineer new features.
- Modeled five regressor models to interpret the importance of features and concluded a decision tree with 82% accuracy.

PUBLICATIONS

- Y. Liang et al., "An Interpretable Image Denoising Framework Via Dual Disentangled Representation Learning," in *IEEE Transactions on Intelligent Vehicles*, [doi: 10.1109/TIV.2023.3331017](https://doi.org/10.1109/TIV.2023.3331017).
- *Survey of Receipt Identification and Classification using Machine Learning. Our Heritage, Vol-68-Issue-15, January 2020, 278-285*