HARSHAL TALELE

📞 +1 (224) 503-7866 |✉ htalele21@gmail.com |🌐 harshaltalele.com |🔗 linkedin.com/in/harshaltalele

EDUCATION

Master of Science (Data Science) — **Aug 2022 – Dec 2023**
University of Rochester, USA — GPA – 3.62/4
- Courses – NLP, Statistical Machine Learning, Time Series Analysis, Computational Statistics, Data Mining
- Graduate Teaching Assistant – Python Analytics at Simon Business School

Bachelor of Engineering (Computer Science) — **Aug 2022 – Jun 2020**
Savitribai Phule Pune University, India — GPA – 8.49/10

SKILLS

Programming & Database Languages – Python, R, PySpark, SQL, MySQL, PostreSQL, NoSQL, MongoDB, Google Firebase
Software Tools & Technologies – Databricks, Advanced MS Excel, Tableau, Alteryx, MLFlow, Git
Technical Skills – Machine Learning, Natural Language Processing, Deep Learning, Hypothesis Testing, Statistical Analysis, A/B Testing

PROFESSIONAL EXPERIENCE

Data Analyst — **Sep 2020 – Jun 2022**
IQVIA (159 Solutions Inc.) **-** *Healthcare Data Science company*
- Delivered analytics to support the pharmaceutical client's drug **sale of $5.6B** with an **increase of 98%** over the previous year
- Built data pipelines and designed Tableau dashboards to report the COVID-19 epidemic metrics and the EDI 867, EDI 852 sales data
- Collaborated in the migration of cloud database infrastructure from Teradata & Azure to AWS S3 by performing QC and sanity checks
  Skills – Tableau, SQL, Advanced MS Excel, Alteryx, Quality Assurance and Control, Business Acumen, Data Storytelling
  Achievement – Presented with 'Ovation Award' for managing offshore work stream that included organizing
  daily catch ups, client interaction & weekly stakeholder deliverables

Data Science Intern — **Aug 2019 – Feb 2020**
Medialytics - *Data Science driven Ed-Tech developing startup*
- Developed a decision support system for content creation to optimize the digital marketing budgets by predicting trend durations
- Automated data engineering pipelines to store hourly Twitter, YouTube, and Google Trends data in multiple levels of data tables
- Performed exploratory data analysis in R & built classification machine learning models to achieve an **accuracy of 84%**
  Skills – Python, R, PostgreSQL, Machine Learning, Exploratory Data Analysis, Git, Leadership
  Achievement – Spearheaded a team of 8 in the software development phase of the project & planned the
  hand-over to a new team in limited duration of time

PROJECTS

Forecasting Bike Inventory for Citibike | *GitHub Link* — **Apr 2023 – May 2023**
- Developed an application for Citibike having **50k+ daily users** to ensure availability of bikes and empty docks at the station
- Implemented ETL pipeline with Spark streaming to process historical data, creating bronze, silver, and gold data-level delta tables
- Performed EDA, built optimized forecasting model for net bike change, and fine-tuned hyperparameters for enhanced accuracy
  Skills – PySpark, Databricks, MLFlow, Time Series Forecasting, ETL, EDA, Git

Dynamic QA Generator for Research Papers | *GitHub Link* — **Mar 2023 – May 2023**
- Developed a QA model to aid efficient comprehension of research papers by summarizing relevant information in the form of Q&A
- Fine-tuned T5 models on OpenAI-modified QASPER dataset with **1.5k+ papers** for question generation and answer generation task
  Skills – Python, Large Language Models, Data Preparation, OpenAI API

Analyzing Political Interest of Indian American | *GitHub Link* — **Nov 2022 – Dec 2022**
- Analyzed tweets from **2.8k unique** Indian Americans to identify their political trends and biases ahead of 2024 presidential elections
- Designed visuals illustrating prevalent topics and sentiment within the community based on party affiliations and geographic locations
  Skills – Python, Topic modelling, Sentiment Analysis, Tableau, Twitter API

COVID-19 Cases Prediction in Ohio | *Kaggle Competition Link* — **Apr 2023 – May 2023**
- Built ML model to predict COVID-19 cases using Ohio county's time series data and analyzed tweet-based social awareness impact
- Achieved top 30% ranking with an impressive **R2 score of 0.89** in the Kaggle competition
  Skills – Ensemble Learning, Feature Engineering, Model Evaluation

100Ascent | *Mobile App Link* — **Aug 2021 – Dec 2021**
- Designed NoSQL database for the fitness-motivation app facilitating seamless fitness tracking and virtual travel experiences
- Achieved **50% increase in active users** by conducting user behavior analysis and leveraging insights to drive business decisions
  Skills – NoSQL, MS Excel, Data Engineering

EXTRA-CURRICULAR

- Selected among top 10 finalists at the IBM Hackathon 2020
- Managed a freelancing team to deliver 20+ writing projects to client
- Hosted formal & informal events for 500+ students & faculty in Annual College Gatherings
- Developed & hosting a WordPress website for my poems & blogs - Link
- Developed a multi-page responsive website for a fitness motivation company Fitno-D