Hrishikesh Mahesh Telang

telanghrishi@gmail.com | +91 91523 90659 | LinkedIn | GitHub | Tableau Public | Google Scholar

EDUCATION

Syracuse University – ***School of Information Studies (iSchool)*** ***May 2023***
Master of Science in Information Systems, Certificate in Advanced Studies—Data Science
Relevant Research: Curating Virality: Exploring Curated Logics within #BlackLivesMatter on Twitter/X
Research Supervisor: Dr. Jeff Hemsley
Awards & Honors: iSchool Values Merit Award

Mumbai University – ***St. Francis Institute of Technology*** ***Nov 2020***
Bachelor of Engineering in Computer Engineering, General
Thesis: Malaria Parasite Classification using Image Processing and Machine Learning
Advisor: Dr. Kavita Sonawane

RESEARCH EXPERIENCE

Research Assistant – ***St. Francis Institute of Technology – Mumbai, India*** ***Aug 2024 – Present***
- Designed and developed an Applicant Tracking System (ATS) with Google Gemini Pro using Gemini API to provide real-time resume analysis and feedback on strengths, weaknesses and areas for improvement.
- Developed an end-to-end Text-to-SQL application leveraging Google Gemini Pro to convert natural language queries to SQL commands and retrieve data from SQLite database via Streamlit interface for non-technical users.
- Implemented an end-to-end document Q&A chatbot by integrating Google's Gamma model and Groq API's LPU engine to enhance the document performance through embedding techniques.
- Conducted literature review on forensic analysis of fake images using machine learning, deep learning, and generative AI techniques.

Product Analyst – ***Revere VC – San Francisco, CA – Contract*** ***Apr 2024 – May 2024***
- Managed quarterly investment performance data of 120+ financial statements, using Python and AI OCR algorithms to automate extraction from financial statement PDFs for cleaner performance tracking.
- Executed sophisticated data cleaning and management protocols with Python on Google Colab and Excel, ensuring integrity and accuracy.

Research Analyst – ***Syracuse University School of Information Studies – Syracuse, NY*** ***Jan 2024 – June 2024***
- Conducted literature reviews on digital transformation, social media and AI use in content moderation to identify emerging trends and gaps.
- Summarized and synthesized findings from case studies and academic sources for ongoing research.

Data Analyst – ***The Washington Post – Washington, DC*** ***June 2023 – Dec 2023***
- Leveraged Python, AWS Redshift SQL, Google Analytics, and Looker to analyze user behavior and campaign performance across platforms, improving the average article conversion rate by ~ 2%.
- Improved workflow efficiency of data pipelines by 52% by incorporating Airflow DAGs to streamline the process of updating newsletter and campaign lookup tables for dashboard content management.
- Spearheaded construction of robust data pipelines and collaborated with data scientists to craft an uplift model to determine pricing strategy.
- Built Tableau Dashboards to evaluate and compare engagement metrics, subscriber behavior, and promotional impact by analyzing data on engagement, visits, and user activities across sales periods, newsletter campaigns, and promotions to inform strategic decision-making.

Graduate Research Assistant – ***Syracuse University – Syracuse, NY – GitHub Link*** ***Jan 2022 – May 2023***
- Scripted open-source STACKS Collection Tool via Twitter API in Python 3.9 to collect tweets and posts on Black Lives Matter, and the 2022 midterm election campaigns.
- Built a BERT model over containerized Ubuntu servers by deploying Docker and built machine learning pipelines to process the semantic analysis of around 1 million tweets, achieving an F1 score of ~ 65%.
- Developed a speech-to-text model by automating video transcription via Google's speech-to-text API, reducing transcription time by 55%.

Data Analyst – ***Student Outreach and Retention (SOaR) at Syracuse University – Syracuse, NY*** ***Jan 2022 – May 2023***
- Analyzed student data within the early alert advising system to provide actionable insights and feedback via data visualization and reporting.
- Built Tableau dashboards on ~ 100K rows for tracking student behavior by performing data wrangling, identifying trends and behavior on a time-series graph, preparing dashboard metric reports, and troubleshooting system errors.
- Provided 6-year data trends to inform programs to support underrepresented students to complete degree requirements.
- Conducted regression and statistical analysis on student demographic data including race, gender, and citizenship status to determine their impact on course and grade performance, achieving 93.60% accuracy.

Data Analyst Intern – ***Office of Institutional Effectiveness at Syracuse University – Syracuse, NY*** ***May 2022 – Aug 2022***
- Boosted process efficiency to 85% by implementing a Python-based web scraper to extract 5K+ faculty directories from XML into Excel spreadsheets, streamlining program review and maximizing academic success and institutional effectiveness.

- Coordinated with a team of three to perform statistical analysis and use Excel functions to consolidate a university data book of 2K+ courses from multiple sources and collaborated with the board to glean business insights for academic and strategic planning.
- Employed Tableau and Power BI to design informative dashboards and used Excel to perform descriptive statistics on graduation reflection survey data, effectively examining student learning outcomes.

Research Assistant – ***St. Francis Institute of Technology – Mumbai, India*** ***Apr 2020 – Jun 2021***

- Utilized KNN, Random Forest, and SVM algorithm in conjunction with an innovative feature extraction algorithm to enhance the classification of malaria parasitized and uninfected blood smear images, achieving 95% accuracy.
- Implemented feature selection techniques on 96 feature vectors for more accurate classification, improving accuracy to 97% using Python and Excel for analysis and visualization.
- Leveraged Python to evaluate the system using binary and multiclass classification for 1,259 COVID-19/Normal/Viral Pneumonia X-ray images and devised a pseudo-coloring algorithm and Global Histogram Equalization (GHE) function.
- Extracted 96 components from the bins approach feature extraction algorithm, performed feature engineering techniques, and obtained >88% accuracy using boosting classifiers and stacking ensembles.

PUBLICATIONS

- **Telang, H.,** & Sonawane, K. (2023). COVID-19 and malaria parasite detection and classification by bins approach with statistical moments using machine learning. *International Journal of Image, Graphics and Signal Processing*, 15(3), 1–13. https://doi.org/10.5815/ijigsp.2023.03.01
- **Telang, H.,** & Sonawane, K. (2020). Effective performance of bins approach for classification of malaria parasite using machine learning. In *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)* (pp. 427–432). IEEE. https://doi.org/10.1109/ICCCA49541.2020.9250789

SELECT RELEVANT PROJECTS

PDF Document Chat Application Using Gemini API and LangChain ***Project Link; July 2024***

- Engineered an end-to-end PDF document chat application to split text chunks and convert >1 GB of text data into vector embeddings for efficient search and retrieval using Google Gemini API and LangChain.
- Developed a user-friendly interface with Streamlit for real-time PDF document uploading, processing, and querying.
- Implemented advanced vector embedding techniques using Facebook's model and Google Generative AI, transforming text chunks into searchable vectors, and designed custom prompts for detailed, context-aware answers with Google Gemini Pro.

Azure-based Data Pipeline and Power BI Dashboard for NYC Taxi Data Analysis ***Project Link; Oct 2022 – Dec 2022***

- Designed and optimized an end-to-end data pipeline using Azure Synapse Analytics, Azure Data Lake Storage with built-in Serverless SQL Pool to create a Power BI dashboard with KPIs, enabling stakeholders to assess company reporting metrics.
- Implemented a database schema with scripts handling various file formats and partitioned data to extract, transform, and aggregate NYC taxi data into a meaningful dataset of 1099 records.

AWS ETL Data Pipeline on YouTube Data *(Python, AWS)* ***Project Link; Jul 2022 – Aug 2022***

- Built a data lake from scratch by creating an ETL pipeline to transform 500 MB+ JSON and CSV files into Apache Parquets to analyze YouTube user data on views, most-watched channels, likes, and dislikes across several countries.
- Leveraged Python to develop a trigger mechanism to detect new JSON inputs in a raw data S3 bucket using AWS Lambda; executed crawling operations, cataloged data, and stored the clean data in a new S3 bucket using AWS Glue.

AG News Text Classification using BERT *(Python, NLP, Transformers, Deep Learning)* ***Project Link; Aug 2022***

- Implemented a multi-class text classification system with Deep Learning using BERT to classify 100K-row dataset into category labels and enhance semantic parsing.
- Performed exploratory text analysis, preprocessed data using ktrain, and fine-tuned the model with TensorFlow, achieving 95% accuracy.

Job Salary Prediction System *(Python, Tableau, NLP, Machine Learning)* ***Project Link; Mar 2022 – Apr 2022***

- Automated median salary predictions for 100K+ UK job listings using Python by incorporating regression and text mining approaches to streamline salary estimation and reduce job seekers' time and effort in gauging potential earnings.
- Modeled and trained SVM, RF and DT Regressors with R2 Score of >0.85 and Naïve Bayes classification model, procuring 79% accuracy.

Hotel Industry Analysis *(R, Predictive Analysis, Linear Modeling, Machine Learning)* ***Project Link; Oct 2021 – Dec 2021***

- Analyzed and cleaned a dataset of 40K+ bookings using R to give business insights and made recommendations to increase Average Daily Rate (ADR) by ~ 35% and to reduce cancellation rates.
- Conducted statistical analysis, found patterns using exploratory data analysis and association rules mining, and obtained classification accuracies of 84.32% for Logistic Regression, 85.28% for Random Forests, and 81% for Support Vectors.

LEADERSHIP & ACTIVITIES

University Senator and Academic Program Senator, Graduate Student Organization, Syracuse University ***Aug 2021 – May 2023***

- Represented the department as an advisor to approve new curricula and administrative decisions, collaborated with other senators to address campus social issues, and addressed grievances.
- Worked with the Outreach and Civic Engagement Committee to increase student awareness by planning and hosting collaborative events and developing advocacy proposals for the GSO Senate.

Toastmasters International – Vice President Public Relations – ***Mumbai, India*** ***Jan 2021 – Jul 2021***

- Led initiatives to improve visibility and membership acquisition for Mumbai Speakers Toastmasters Club through social media campaigns and distributing promotional content.
- Collaborated with the executive team to organize District Speech Contests, and enhanced the club's engagement with members of other clubs.

Rotaract – Community Service Co-ordinator – ***Mumbai, India*** ***Jul 2020 – June 2021***

- Organized a seminar on 'Period Leave' in collaboration with 23 Rotaract Clubs, featuring speakers from One Future Collective.
- Hosted the IGTV web series 'Sakhi Tales,' addressing global menstrual issues through interviews highlighting personal stories.
- Coordinated a discussion with Sobhan Mukherjee (Kolkata 'Padman') on menstrual activism and LGBTQ+ advocacy, with participation from 15 Rotaract clubs.
- Participated in multiple donation and collection drives (sanitary products, food, clothing) and organized community service initiatives, including yoga sessions, nursing home visits, and education programs through Project 'Aashayein.'

TEDxSFIT – ***Co-organizer – Mumbai, India*** ***Feb 2019 – Aug 2019***

- Spearheaded and supervised the inaugural TEDxSFIT event as a founding and core committee member.
- Directed a team of four in curating and drafting compelling speeches, sponsorship proposals, and various event materials.
- Managed communication, collaborated with officials, and oversaw various aspects of event planning to ensure a smooth program flow.

AWARDS & HONORS

- Best Member Award – Rotaract Club of Bombay Mahakali Heights (Rotaract 3141)
- WOW! What Outstanding Work Award – Rotaract 3141
- Awarded Third Place for the Research Project "Malaria Parasite Classification using Image Processing and Machine Learning" - Department of Computer Engineering, St. Francis Institute of Technology

CERTIFICATIONS

MOS: Excel Associate (Microsoft 365 Apps, Office 2019), MOS: Microsoft Excel Expert (Office 2019)

TECHNICAL SKILLS

Data Science Tools: Python (Numpy, Pandas, Matplotlib, Scikit-learn, NLTK, OpenCV, Scikit-image), R, C/C++, Airflow
Database Management Tools: MySQL**,** SQL Server, MongoDB, MS Access, Snowflake, ETL
BI & Analytics Tools: Tableau, Google Analytics, Power BI, Looker
Application Software: R Studio, Azure Data Studio, Jira, Confluence, Jupyter Notebook, Hadoop, MS Office, GitHub (Version Control), Google Gemini Vision
Cloud Services: AWS Services (S3, Lambda, QuickSight, Athena, Glue, Identity & Access Management), Microsoft Azure, Docker