Aditya Gaydhani

612-812-7744 • aditya.gaydhani@gmail.com • Minneapolis, MN • adityagaydhani.github.io • linkedin.com/in/adityagaydhani • github.com/adityagaydhani

EDUCATION

University of Minnesota – Twin Cities, MN — **Expected Graduation: December 2020**

Master of Science in **Data Science** | Minor in **Statistics** — GPA: 3.80/4.00

- **Coursework** – Data Mining, Computer Vision, Artificial Intelligence, Big Data Engineering & Analytics, Distributed Systems, AI for Earth *(Deep Learning in GIS and Remote Sensing)*, Statistics I/II, Applied Regression, Capstone Project: Dialogue Systems.

University of Pune, India — **June 2018**

Bachelor of Engineering in **Computer Engineering** — GPA: 3.55/4.00

- **Selected Coursework** – Algorithms, Data Structures, Databases, Software Engineering, Operating Systems, Computer Networks.
- **Honors and Awards** – Top 60 nationwide *(~11,000 participants)* in Computer Society of India '17 Contest, based on Algorithms.

PUBLICATIONS

- **A. Gaydhani**, R. Finzel, S, Dufresne, S. Pakhomov, M. Gini
  Conversational Agent for Daily Living Assessment Coaching
  AI4Function Workshop, **International Joint Conference on Artificial Intelligence (IJCAI), 2020** [preprint] [demo]
- **A. Gaydhani**, V. Doma, S. Kendre, L. Bhagwat
  Detecting Hate Speech and Offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach
  IEEE 8th International Advance Computing Conference (IACC), 2018 – Poster Presentation arXiv:1809.08651

TECHNICAL SKILLS

- **Languages** – Python *(strong)*, C, C++, SQL, R, Java
- **Frameworks** – TensorFlow, Keras, Scikit-learn, Pandas, NumPy, NLTK, SpaCy, Matplotlib, Seaborn, Flask
- **Other** – Git, MongoDB, Apache Hive, Apache Spark, Apache Kafka

WORK EXPERIENCE

Graduate Research Assistant — **May 2019 – Present**

Department of Pharmaceutical Care & Health Systems, University of Minnesota — Minneapolis, MN

- Develop a tutoring **dialogue system** for Minnesota Department of Human Services to correctly assess people during interviews.
- Designed pipelines to collect real-world dialogues from assessors; improved data quality by using preprocessing techniques.
- Built a **Bi-LSTM** neural network with **GloVe embeddings** for domain and intent classification; achieved **~89%** and **~94%** f-scores.
- Facilitated generation of **personality-based** profiles of dialogue agent by utilizing de-identified historical evaluation data.
- Leveraged transformer-based models like **BERT** and **DialoGPT** for Named Entity Recognition and Natural Language Generation.
- Extended the functionality of MindMeld, an open-source conversational AI platform, by integrating these deep learning models.

Software Developer — **June 2019 – August 2019**

Department of Forest Resources, University of Minnesota — Minneapolis, MN

- Developed an **R package** end-to-end, to compute functional biodiversity indices through graph algorithms using forest data.
- Collaborated within a **cross-functional** team to devise software requirements and design; followed **agile** methodology.
- Solved performance bottlenecks by integrating speedy compiled C++ code, utilizing parallelization, and using robust non-linear optimization techniques; reduced computational time by **~66%**.

ACADEMIC PROJECTS

Extracting Water Bodies from Satellite Images using Deep Learning — **Python, TensorFlow**

- Built a robust **semantic segmentation** model to improve inference time to map water, as compared to rule-based algorithms.
- Prepared a labeled image dataset; efficiently loaded **24GB** of data for training on limited resources by using data generators.
- Reduced memory footprint by **~83%** by adapting the model architecture for training high resolution multi-band satellite images.
- Outperformed the **FCN** and **UNet** models by an f-score of **~13%** and **~6%**; evaluated the model for spatial heterogeneity on earth.

Big Data Analysis to Understand Operational Efficiency of Flight Carriers — **Spark, Hive, Zeppelin, Kafka, AWS**

- Constructed a **data pipeline** to load, transform, and aggregate **~5M records** using Apache Kafka, deployed on an AWS instance.
- Formulated and addressed business questions using Hive; improved the retrieval time by optimizing the queries using **caching**.
- Predicted flight cancellations using Spark MLlib with **~86%** accuracy; communicated data-driven insights through visualizations.

Toxic Language Detection on Twitter — **Python, Scikit-learn, Pandas, NLTK**

- Enriched existing dataset by retrieving over **70K tweets** using Twitter API; performed **feature engineering** to extract tf-idf values.
- Implemented and evaluated Naive Bayes, SVM & Logistic Regression models; tuned hyperparameters to achieve **~96%** f-score.
- Interfaced the classifier with Twitter using **Flask** and Twitter API to create a real-time web application for detecting toxic tweets.

Scene Recognition using Visual Bag-of-Features from Images — **MATLAB**

- Designed and constructed a set of visual recognition systems using KNN and SVM that classify images into scene categories.
- Improved accuracy by **~18%** as compared to a baseline model, by leveraging robust features from images such as **Dense SIFT**.
- Performed dimensionality reduction by **clustering** features using K-means; constructed visual bag-of-features per image.