Aditya Gaydhani

612.812.7744 | gaydh001@umn.edu | Minneapolis, MN | adityagaydhani.github.io | linkedin.com/in/adityagaydhani | github.com/adityagaydhani

SUMMARY

Seeking a full-time Software Development Engineer or Data Scientist position starting June 2020.

EDUCATION

University of Minnesota – Twin Cities   **MN, USA**
Master of Science in *Data Science* | Minor in *Statistics*   Expected Graduation: May 2020
- **Relevant Coursework** – Applied Regression Analysis, Time Series Analysis, Big Data Engineering & Analytics, Artificial Intelligence, Computer Vision, Data Mining, Machine Learning

University of Pune   **Pune, India**
Bachelor of Engineering in *Computer Engineering*   June 2018
- **Relevant Coursework** – Data Structures, Algorithms, High Performance Computing, Cloud Computing, Distributed Systems
- **Honors & Awards** – Achieved a position among top 60 nationwide out of **~11,000 participants** in Computer Society of India Programming Contest 2017, a national level problem solving contest based on Algorithms and Data Structures

TECHNICAL SKILLS

- **Strong** – Python *(TensorFlow, PyTorch, Keras, Scikit-learn, Beautiful Soup, Flask, Plotly)*, R *(Tidyverse)*, C, C++, SQL, MongoDB
- **Familiar** – Java, MATLAB, Hadoop, Hive, Spark, Pig, Kafka
- **Tools** – Git, Tableau, Weka, LaTeX, Jupyter Notebook, RStudio
- **Other** – Machine Learning, Natural Language Processing, Data Engineering, ETL, Computer Vision, R Package Development, Web Scraping, Data Visualization

WORK EXPERIENCE

Pharmaceutical Care & Health Systems, University of Minnesota   **MN, USA**
Graduate Research Assistant – Natural Language Processing – Dr. Serguei Pakhomov   May 2019 – Present
- Developing a **dialogue system** for Minnesota Department of Human Services, designed to train assessors to conduct interviews.
- Generating synthetic profiles, that represent people with certain disabilities, using **GANs** based on the past evaluation data.
- Working on Topic Classification and Semantic Role Labeling using AllenNLP for the NLU and NLG components of the system.

Department of Forest Resources, University of Minnesota   **MN, USA**
Software Developer – R Package Development   June 2019 – Aug. 2019
- Developed an **R package** end-to-end, to compute functional biodiversity indices using forest data (to be published on CRAN).
- Solved performance bottlenecks by integrating speedy compiled C++ code, utilizing parallelization, and using robust non-linear optimization techniques like IPOPT, which together optimized the code performance by **six times**.

PUBLICATIONS

- Detecting Hate Speech and Offensive Language on Twitter using Machine Learning: An N-gram and TFIDF based Approach
  A. Gaydhani, V. Doma, S. Kendre, L. Bhagwat; **IEEE IACC 2018** (Poster Presentation) arXiv:1809.08651 [cs.CL]

ACADEMIC PROJECTS

Toxic Language Detection on Twitter   *Tech Stack: Python, Scikit-learn, Pandas, NLTK*
- Developed an application in Python to automatically detect hate speech on Twitter, achieving up to **96% F-score**.
- Pulled over **70K tweets** using Twitter API and performed data preprocessing and cleaning using **NLTK** to improve data quality.
- Performed feature engineering to extract TFIDF values & evaluated performance of Naive Bayes, SVM, and Logistic Regression.
- Interfaced the application with Twitter using **Flask** and Twitter API to detect tweets containing toxic language in **real-time**.

Analyzing Flights and Airport Data using Hadoop and Spark   *Tech Stack: Hadoop, Spark, Hive, Zeppelin*
- Collaborated with a team of four to analyzed over **5M instances** of Flights and Airport data using Hive and Spark to understand statistics of flight carriers and airports and performed **interactive data analytics** using Apache Zeppelin.
- Designed and constructed a machine learning model using **Spark MLlib** to predict flight cancellations with **~86% accuracy**.

Big Data Analysis on Urban Dictionary Data   *Tech Stack: MongoDB, Elasticsearch, Kafka*
- Engineered a Big Data System to analyze over **2.9M** JSON documents on Urban Dictionary using NoSQL technologies.
- Established **data pipelines** using Apache Kafka to **automate** the process of validating data from Urban Dictionary API.
- Optimized the queries using **caching** and performed data visualization to gain data-driven insights about Urban Dictionary.

Multi-frame Face Tracking   *Tech Stack: MATLAB*
- Devised a robust face tracker by mapping SIFT features to the template image and improving their quality using RANSAC.
- Increased the performance by **3x** using inverse compositional alignment & refining each frame to effectively track the template.

Stereo Reconstruction   *Tech Stack: MATLAB*
- Implemented stereo reconstruction by extracting key points and computing fundamental matrix to locate the epipole.
- Performed triangulation, pose disambiguation, and stereo matching to get a disparity map used for computing image depth.