

GAURAV MUDBHATKAL

gmudbhat@uwaterloo.ca ♦ (519) 729-2861 ♦ [LinkedIn](#) ♦ [GitHub](#) ♦ University of Waterloo

EDUCATION

Honors, Bachelor of Applied Sciences in Management Engineering | University of Waterloo **Apr. 2022**

Relevant Courses: Data Mining, Software Engineering, Optimization, Statistics, Human Computer Interaction

SUMMARY OF SKILLS

- **Programming Languages:** Python, Scala, Java, SQL, HTML/CSS, JavaScript, C++/C#
- **Big Data & ML:** Spark, TensorFlow, Scikit-learn, NumPy, pandas, Keras, NLTK, Hadoop, MongoDB
- **Technologies & Frameworks:** Azure, Git, Linux, Anaconda, Jenkins, Docker, AWS, jFrog, Node, React
- **Project Management:** Agile, Problem Analysis, JIRA, Confluence, Notion, ServiceNow, Trello

WORK EXPERIENCE

Machine Learning Intern | Thomson Reuters – Legal Tracker **Sept. 2021 – Dec. 2021** | Toronto, Canada

- Tuned and evaluated **document classification** models after conducting surveys across various departments
- Orchestrated model deployment pipelines to integrate machine learning services with **Azure DevOps**

Data Engineer Co-op | Manulife – Canada Data Office **Jan. 2021 – Apr. 2021** | Remote, Canada

- Simplified analytics in the cloud by writing **Spark** code to facilitate warehousing of data in **Azure Data Lake**
- Developed templates for data science pipelines in **Azure Synapse** to support future machine learning projects
- Coordinated the migration of ~30 repositories to **GitHub** with ServiceNow, JIRA, and Confluence

Data Engineering Intern | KPMG – Lighthouse Exponential **Jul. 2020 – Aug. 2020** | Remote, Canada

- Delivered subsidies to clients affected by COVID-19 as part of the Canada Emergency Wage Subsidy program
- Calculated upwards of \$2M in eligible subsidies by wrangling client payroll data using workflows in **Alteryx**

Data Scientist | RBC - Global AML Risk Analytics **Sep. 2019 – Dec. 2019** | Toronto, Canada

- Generated **Doc2Vec** embeddings for ~10,000 documents after preprocessing & tokenizing text using **NLTK**
- Achieved 96% accuracy & 82% precision through a neural network-based multi-label text classifier with **Keras**
- Automated python package deployment to the **jFrog Artifactory** by integrating with **GitHub** & **Jenkins**

Data Analytics Co-op | BlackBerry Messenger **Jan. 2019 – Apr. 2019** | Mississauga, Canada

- Streamlined data processing by writing Apache **Spark** jobs with **Scala** & warehousing with **BigQuery**
- Optimized the data team's ETL pipeline by scheduling Spark jobs using Apache **Airflow**
- Investigated spikes in user data by visualizing trends with **Tableau** & Apache **Zeppelin**

PROJECTS & EXTRACURRICULARS

COVID-19 Twitter Sentiment Analysis | Data Mining Course Project | [GitHub](#) **Dec. 2020**

- Performed **text mining** on ~12M tweets related to COVID-19 using Natural Language Processing techniques
- Applied **classification** algorithms to detect sentiment, analyzed reaction to the pandemic through **clustering**

Enterprise Co-op Student | Conrad School of Entrepreneurship & Technology **May. 2020 – Jul. 2020**

- Analyzed & conducted research on the storage & analysis of digital evidence data by Law Enforcement
- Pitched research for a chance to win \$15,000 in funding through the Problem Pitch Competition

Infrastructure Team Member | [WATonomous](#) – Self Driving Car Team **May. 2019 – Aug. 2019**

- Assisted in containerization of the team's autonomous vehicle developer environment using **Docker**