

# Siddardha Kaja

Toronto, ON • [siddhardhakaja007@gmail.com](mailto:siddhardhakaja007@gmail.com) • (437) 435-6171 • <https://github.com/siddardhakaja>

## EDUCATION

---

### ACADIA UNIVERSITY

Wolfville, NS

*Master of Science in Computer Science with Co-op*

2019-2021

- **GPA 3.67/4.00**
- **Thesis :** Authored a novel Backup One Hop Routing Optimization framework combining advanced techniques and deep learning models to enhance data packet efficiency on high-bandwidth backbone routers - <https://siddardhakaja.web.app/assets/doc/thesis.pdf>

## PUBLICATIONS

---

- [Acknowledgment scheme using cloud for node networks with energy-aware hybrid scheduling strategy](#) - Journal of Ambient Intelligence and Humanized Computing (2020)
- [Long Short-Term Memory Approach for Routing Optimization in Cloud ACKnowledgement Scheme for Node Network](#) - The 12th conference on Ambient systems, Networks and Technologies (ANT 2021)

## PRESENTATIONS

---

- Emergency Rescue Tool for Ferries using Computer Vision - DeepSense Ocean of Data Challenge : The Bedford Ferry, Halifax, NS (2021)

## PROFESSIONAL & RESEARCH EXPERIENCE

---

### AIR CANADA

Toronto, ON

Data/ML Engineer

2022-Present

- Introduced LLM-powered contextual YouTube comment sentiment analysis using Pyspark in Databricks and built a Power BI dashboard to deliver actionable brand insights
- Migrated an everyday data cleansing task that cleans about 70 M records from Databricks to Snowflake, increasing efficiency by 20% and speed by 40% through data movement reduction and using LLM and RAG to further enhance the cleansing operation
- Built API connectors in Databricks and Dataiku to streamline data ingestion from Google Marketing Platforms and Facebook Ads Manager, empowering marketing analytics teams to independently fetch reports and increasing team productivity by 10%
- Created a reusable Databricks framework for data decryption, validation, and transformation, reducing the new ETL pipeline development time by 80%
- Developed ADF pipelines to ingest and transform diverse campaign data from Google Cloud Storage into Azure Data Lake, optimizing data for downstream consumption in Databricks and enhancing marketing attribution modeling
- Designed automated failover and recovery mechanisms for Azure Data Factory pipelines processing critical airline data, ensuring uninterrupted data flow for downstream consumption by multiple teams for post travel reporting, increasing overall team efficiency by 5%

### ACADIA INSTITUTE FOR DATA ANALYTICS

Wolfville, NS

Research Assistant

2020-2021

- Architected data cleaning pipeline for large-scale textual datasets using Google Colab, enhancing data quality and efficiency for subsequent NLP modeling. Created LSTM, Naive Bayes and Decision Tree models to address a complex NLP classification problem
- Extracted and analyzed complex mass spectrometry data to identify key chemical components. Developed and trained machine learning and statistical models utilizing the cleaned data to accurately predict the overall chemical composition of samples

- Conducted thorough requirements analysis to create detailed documentation, effectively communicating project scope and target to stakeholders
- Performed in-depth analysis of household power consumption data, utilizing Matplotlib and Seaborn to visualize and identify consumption patterns, anomalies, and potential energy-saving opportunities

## COMMUNITY INVOLVEMENT

---

- Acadia Robotics, Student Volunteer (2019-2020)
- Acadia International Student Ambassador (2019-2020)

## REFERENCES

---

- **Dr. Elhadi Shakshuki**  
Professor of Computer Science at Acadia University, **Email :** [elhadi.shakshuki@acadiau.ca](mailto:elhadi.shakshuki@acadiau.ca)
- **Dr. Danny Silver**  
Professor of Computer Science at Acadia University, **Email :** [danny.silver@acadiau.ca](mailto:danny.silver@acadiau.ca)
- **Dr. Haroon Malik**  
Asst. Professor at Computer Science at Marshall University, **Email :** [malikh@marshall.edu](mailto:malikh@marshall.edu)
- **Naveen Rajan**  
Manager, Data and Analytics at Air Canada, Email : [naveen.rajan@aircanada.ca](mailto:naveen.rajan@aircanada.ca)

## ADDITIONAL INFORMATION

---

- **Technical Skills:** Python, Pyspark, SQL, Snowflake, Snowpark, Streamlit, Dataiku, Databricks, Azure Data Factory (ADF), Azure Data Lake (ADLS), Power BI, Git, Confluence, Machine Learning, NLP, Gen AI/LLM, Prompt Engineering, RAG, ETL modeling
- **Soft Skills:** Problem Solving, Teamwork, Communication Skills, Time Management, Adaptability, Collaboration
- **Languages:** Fluent in English and Telugu, Conversational Proficiency in Hindi