



Abhijeet Singh

Data Engineer II - Bell Canada

📍 Toronto, ON 📞 782-446-6007 ✉ abhijeetsingh0727@gmail.com **in** [Linkedin](#)  [Github](#)  [Website](#)

Highlights

- Skilled Data Engineer II and ML Developer with 5+ years of expertise in building and optimizing end-to-end Data and ML pipelines
- Highly proficient in developing and deploying ELT/ETL data pipelines, data modelling & warehousing, classification & prediction modelling using NLP, LLMs and statistical methods with 7+ years of research and industrial experience
- Demonstrated strong leadership, representation, and teamwork by collaborating with diverse stakeholders, eagerly tackling new challenges and adapting flexibly to evolving project demands
- AWS Cloud Practitioner, Databricks Applied data Science, Databricks Apache Spark SQL for Data Analyst, Bell Excellence and Innovation Award 2022,2024

Relevant Work Experience

Data Engineer II

Jan. 2024 - Present

Bell Canada, Toronto, ON

- Led design and automation of **10+ data pipelines** ingesting millions of records daily from SAP and legacy systems to GCP, using **Cloud Composer - Airflow, Cloud Function, Data Fusion, Big Query** optimizing **ELT** workflows on GCP for market segmentation, with **40%** faster queries, **25%** cost reduction, and **99.9%** uptime
- Designed and implemented **scalable and optimized data models** with robust **backend APIs** to efficiently **manage millions** of sales-representative records, orchestrating end-to-end data pipelines on **GCP** with Cloud SQL, Data Fusion, and **Cloud Composer**
- Contributed to building and orchestrating **telecom orders ML pipelines** on **Vertex AI** to assign customer segmentation with **88%+** accuracy using **XGBoost**. by utilizing **Cloud Composer, Pub/Sub, BigQuery**, and other **GCP** services
- Built automated **Azure** data pipeline using Data Factory, PySpark, and Synapse for timesheet reconciliation, processing 200K+ records daily, reducing errors by 30% and boosting performance by 40%

Data Engineer I

Jan. 2021 - Dec 2023

Bell Canada, Halifax, NS

- Designed and deployed an end-to-end **fraud detection** pipeline on Azure using Data Factory, Synapse, and **Azure ML** by employing **neural networks**, achieving **92% F1 score** and **\$1M+** savings
- Contributed to nine data science pipelines on **Vertex AI** for customer churn prediction with **87%** accuracy, using Cloud Composer, Pub/Sub, and BigQuery
- Built an automated machine learning pipeline for finance revenue forecasting on **Alteryx**, reducing forecast error **MAPE** by **20%** and manual effort by **40%**
- Gained a deep understanding of **business values** and **agile delivery model** through close collaboration with the business and tech team

Data Scientist and ML Researcher

Jan. 2020 – Jan. 2021

Dalhousie University, Halifax, NS

- Developed an **NLP-driven** binary classification model using **TF-IDF** and **Word2Vec** to identify duplicate Quora questions, **84% precision**, and **90% recall**

- Designed and implemented a **sentiment analysis** pipeline for online reviews using **SpaCy-based** feature engineering and cross-validation, achieving a **90.08% F1 score** with an **SVM model**

Logistics Data Analyst

Sep. 2018 – July. 2019

Star Cement, Guwahati, India

- Handled data cleaning, missing value treatment, and preparation for analysis with full process documentation to support logistics insights
- Created visualizations using Python and Tableau to present user stories and drive data-informed decisions in logistics operations

Education

Masters, Computer Science (ECommerce)

Sept. 2019 – May. 2021

Dalhousie University

Halifax, NS

B.Tech. in Computer Science and Engineering

Apr. 2014 – Aug. 2018

AKTU University

Lucknow, India

Technical & Soft Skills

Languages: Python, Java, C++, SQL, C#, JavaScript, PHP, HTML/CSS, PySpark

Tools: GCP, Azure, Linux, Git, Terraform, Docker, Shell, Bash, ROS, Cognos, Apache Airflow

Visualizations: Tableau, Matplotlib, Seaborn, Plotly, Microsoft Power BI, LIME, Streamlit

NLP: NLTK, Spacy, Gensim, Hugging Face, Stanza, LangChain, DeepSpeed, PaddlePaddle

Libraries: Tensorflow, Pytorch, Keras, Sklearn, OpenCV, YOLO, Pandas, Numpy

Database: MySQL, Microsoft SQL Server, SQLite, PostgreSQL

Soft Skills: Management & Leadership, Critical Thinking, Communication, Work Ethic, Time Management, Creativity

Professional Contributions

- End-to-End Data Engineering: Automating Cryptocurrency Data Pipelines with GCP, Medium. Built a serverless ETL pipeline using Cloud Functions, Cloud Composer, and BigQuery to automate crypto data ingestion and reporting. [Article Link](#)
- GitHub Archive: Data Engineering and Analytics Project, Medium Built a scalable GCP data pipeline using Terraform, Cloud Storage, BigQuery, and Cloud Composer to automate GitHub event analytics and infrastructure provisioning. [Article Link](#)