

# Md Fahimul Kabir Chowdhury

Denton, TX | +1 (940) 364-9486 | [mdfahimulkabirchowdhury@my.unt.edu](mailto:mdfahimulkabirchowdhury@my.unt.edu) | [LinkedIn](#) | [Github](#) | [Portfolio](#)

## EDUCATION

### University of North Texas

Master's in Data Engineering | CGPA: 3.66/4.00

Denton, TX

Aug. 2024 – May 2026

### University of Science and Technology Chittagong

Bachelor of Science in Computer Science and Engineering | CGPA: 3.58/4.00

Chittagong, Bangladesh

Jan. 2019 – May 2023

## EXPERIENCE

### Graduate Research Assistant

University of North Texas

Nov. 2024 – Present

Denton, TX

- Researched transactional behavior analysis using large-scale datasets.
- Designed and tested CNN and transformer models in Python and PyTorch.
- Mentored 150+ undergraduate students in courses on data mining, machine learning, and data engineering.

### EdTech Content Creator

YouTube

July 2020 – Present

Dhaka, Bangladesh

- Created a YouTube channel on programming and software topics (150K+ subs, 10M+ views).
- Developed and presented technical content, demonstrating strong communication and presentation skills.

### Data Engineering Intern

TigerIT

June 2022 – Sept. 2022

Dhaka, Bangladesh

- Designed and optimized a scalable ETL data pipeline using Apache Spark and AWS S3.
- Boosted data availability and cut processing time by 30%.

## PROJECTS

### Credit Card Fraud Detection Pipeline | PySpark, Spark Streaming, Spark MLlib, AWS | [Docs](#)

2025

- Build a real-time fraud detection pipeline using PySpark and Spark Streaming on distributed data.
- Engineered features and trained classification models with Spark MLlib on anonymized transaction data.
- Deployed pipeline components on AWS for scalability; achieved 99.99% accuracy and 0.999 AUC-ROC score.

### California Housing Price Prediction | Python, Pandas, Scikit-learn, Matplotlib | [Docs](#)

2025

- Modeled California housing prices using Scikit-learn with data preprocessing and visualization.
- Conducted data cleaning, feature engineering, and model evaluation with MAE and RMSE metrics.
- Built data pipelines in Python to automate preprocessing and generate real estate value forecasts.

### End-to-End eCommerce Data Pipeline & Forecasting | Hadoop, PySpark, Hive, Tableau | [Docs](#)

2024

- Designed and implemented an end-to-end ETL pipeline for retail demand forecasting and business intelligence.
- Ingested eCommerce data into AWS S3, transformed via PySpark on EMR, and queried with Hive.
- Delivered interactive dashboards via Tableau for tracking sales KPIs and optimizing inventory planning.

## TECHNICAL SKILLS

**Languages:** Python, Java, SQL, JavaScript

**Data Tools:** PySpark, Hadoop, Kafka, Airflow, Pandas

**Databases:** MySQL, PostgreSQL, MongoDB, Snowflake

**Cloud Platforms:** AWS, GCP, Azure

**Infrastructure:** Docker, Kubernetes, CI/CD Pipelines

**AI/ML:** Scikit-learn, TensorFlow, Keras, PyTorch

**Frameworks:** React, Django, Node.js, Flask

## AWARDS AND RECOGNITION

Selected for Top 50 Teams in Bangladesh, NASA Space App Challenge.

YouTube Creator Award, 100K Subscribers Milestone.

Best Thesis Presenter, University of Science and Technology Chittagong.

## CERTIFICATIONS

Exploratory Data Analysis for Machine Learning | IBM | [Coursera](#)

Introduction to TensorFlow for AI, ML, and DL | DeepLearning.AI | [Coursera](#)

Introduction to Applied Machine Learning | Alberta Machine Intelligence Institute | [Coursera](#)