Thuhin Khanna

Montreal, QC
thuhinkhanna02@gmail.com | +1 (438) 921-9035
linkedin.com/in/thuhinkhanna | github.com/thuhinkhanna

Professional Summary

Analytical and detail-oriented Data Scientist with a strong foundation in machine learning, NLP, and data engineering. Skilled at designing scalable data pipelines, training and evaluating ML models, and translating complex data into actionable insights. Experienced with Python, SQL, Power BI, and cloud platforms (Azure, GCP). Passionate about applying data-driven methods to solve real-world problems and optimize decision-making.

Education

Master of Applied Computer Science – Concordia University, Montreal, QC (June 2025) | GPA: 3.45/4.30
Bachelor of Computer Science and Engineering – Anna University, Chennai, India (May 2023) | GPA: 9.24/10.0
Relevant Coursework: Pattern Recognition, Applied AI, Natural Language Analysis, Database Systems.

Technical Skills

Core Competencies: Machine Learning, Natural Language Processing, Predictive Modeling, Statistical Analysis, Data Visualization, A/B Testing.
Programming Languages: Python, SQL, Java, C++, C, Bash Script.
Libraries & Frameworks: PyTorch, Tensorflow, Scikit-Learn, NLTK, Matplotlib, Pandas, NumPy.
Tools & Platforms: Git, Visual Studio Code, Power BI, Jupyter Notebooks, MS Excel.
Cloud & Infrastructure: Microsoft Azure, Google Cloud Platform (GCP), Kubernetes, Linux.
Data Engineering: ETL Pipelines, Data Cleaning, Feature Engineering, Data Wrangling.

Projects

Public Perception Analysis of World Leaders | Python, PyTorch, MongoDB (June 2022)
- Engineered a large-scale Reddit crawler to collect 1GB+ of text data across 50 world leaders.
- Built an end-to-end NLP pipeline for sentiment classification using BERT, achieving 98% accuracy.
- Processed and structured 40,000+ sentences with advanced data cleaning and normalization.
- Visualized sentiment distributions and leader popularity trends using Matplotlib and dashboards.

Real-Time Censorship Detection using Deep Learning | Python, NLTK, TensorFlow, Flask (April 2023)
- Developed and trained multi-class text and image classification models for NSFW detection, exceeding 90% accuracy.
- Curated a 10,000+ labeled dataset for supervised learning and deployed models to cloud with Flask API.
- Created a browser extension that redacts inappropriate content based on model predictions.

Rule-Based Parsing of Patient Information Leaflets | Python, NLTK, BeautifulSoup (Dec 2024)
- Built an ETL pipeline to extract and preprocess medical text data.
- Implemented a custom NER and grammar-based parser, increasing parsing coverage from 22% to 78%.
- Extracted and structured 500+ sentences using BeautifulSoup and RegEx.

Experience

Data Analytics Intern — Ziroh Labs, Bangalore, India (Sept 2021 – Feb 2022)
- Developed and automated ETL pipelines to extract and transform data from SQL stores.
- Designed interactive Power BI dashboards to monitor KPIs and performance metrics.
- Collaborated with cross-functional teams to convert analytical findings into actionable strategies.

SQL Developer Intern — Bharthi AXA Insurance, Chennai, India (May 2022 – Dec 2022)
- Developed a Proof of Concept integrating Microsoft SQL Server with a Django REST backend.
- Deployed and optimized SQL Server in production to improve data reliability and query performance.
- Built API-driven data pipelines for real-time analytics and decision support.

Leadership & Extracurricular

Editor — Texplore Monthly Campus Magazine (Aug 2020 – Dec 2021)
- Authored and edited technical articles for a 150+ member community, enhancing engagement and visibility.