

Gurjeet Kaur

 New York  +1 (551) 804-2930  gk2845@nyu.edu  gurjeetkaur11  gurjeetkaur0005

Education

New York University

M.S. in Computer Science — GPA: 3.89/4.0

Sep 2025 – May 2027

New York, USA

Courses: Big Data, Machine Learning, Foundations of Data Science, Design and Analysis of Algorithms, Information Visualization, Cloud Computing, Open Source / Professional Software Development

Thapar Institute of Engineering and Technology

M.Sc. Mathematics and Computing — GPA: 3.79/4.0

Aug 2022 – May 2024

India

Courses: Operating Systems, Computer Networks, Data Analysis and Visualization, Probability and Statistics, Mathematical Programming

Panjab University

B.Sc. Mathematics, Physics, and Chemistry — GPA: 3.31/4.0

Jul 2019 – May 2022

India

Courses: Real Analysis, Probability, Calculus, Statics, Modern Algebra, Matrices

Technical Skills

Core Programming: Python (advanced), C++, SQL, MATLAB

ML Frameworks: Transformers, CNN, RNN, PyTorch, TensorFlow, Scikit-learn, Predictive Modeling, Regression, Clustering

Data Science & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly, D3.js, SVG, Tableau, Data Analytics, GitHub

Big Data & Scalable Systems: Apache Spark (PySpark), Hadoop (HDFS, MapReduce), Dask, MongoDB, Hive

Databases & Backend: PostgreSQL, SQLite, REST APIs, Flask

Work Experience

iEminence Inc.

Feb 2025 – Jul 2025

IT Intern

- Built a financial dashboard using Python, Streamlit, Pandas, and Matplotlib, enabling real-time profit loss summaries, cash flow analysis, and interactive visualizations for stakeholders.
- Automated complex Excel workflows, standardized multi-sheet financial datasets, and generated dynamic, downloadable reports with charts and key insights using XlsxWriter.
- Led the project end-to-end, including requirement gathering, development, testing, and deployment, improving reporting efficiency by around 30%.
- Enhanced data accuracy and consistency by implementing robust preprocessing pipelines and validation checks across all financial data sources.
- Streamlined decision-making for finance teams by integrating actionable metrics and visual analytics into the dashboard.

Research & Projects

Scaling Laws in Symbolic Music Modeling

- Analyzed scaling behavior of Transformer and LSTM language models using 178K MIDI files, constructing a 143M-token ABC corpus with a 1,734-token note-level vocabulary.
- Trained models from Tiny (~3M) to XL (~100M+) parameters in PyTorch; demonstrated power-law scaling, achieving 2.22 test loss, 9.20 perplexity, and 100% syntactically valid generation.

Online Retail Market Basket Analysis

- Applied PySpark FP-Growth on 1.1M+ transactions to mine association rules for cross-selling.
- Identified 850+ frequent itemsets and 240+ high-confidence rules, evaluated using support, confidence, and lift.

Heart Disease Prediction

- Built a robust heart disease prediction system using an ensemble of Random Forest and CNN, optimized with Genetic Algorithm and Cuckoo Search Optimization.
- Handled class imbalance with SMOTE and improved data quality via preprocessing (normalization, filtering, outlier detection).
- Achieved 95% accuracy on the UCI Heart Disease Dataset, showcasing AI-assisted clinical decision support.

Thesis Research

A Study of Bilevel Integer Quadratic Programming Problems

- Developed a generalized MATLAB algorithm for quadratic and linear bilevel integer programming, addressing limitations of existing BIQP methods and validating correctness via computational experiments.
- Applied the algorithm to real-world scenarios in energy markets, transportation networks, supply chains, and facility location planning, demonstrating practical impact.