Kavya Pratik Mehta
Boston, MA • (973) 462-4828 • mehta.kavy@northeastern.edu • [LinkedIn](#) • [GitHub](#)

EDUCATION

Northeastern University - Boston, MA   **Expected May 2026**
Master of Science in Information Systems
Relevant Courses: Data Science Engineering, Program Structures Algorithms, Cloud Computing, Web Development, Application Engineering

St Francis Institute of Technology - Mumbai   **Dec 2022 - May 2024**
Bachelors in Electronics and Telecommunication
Relevant Courses: Database Management System, Cloud Computing and Security, Data Compression and Cryptography

TECHNICAL SKILLS

Programming: Python, SQL, PySpark, R, JavaScript
Data Engineering: ETL/ELT Pipelines, Data Modeling, Data Quality & Validation, Databricks, Alteryx
Cloud & AWS: AWS, Azure, Terraform, CI/CD
Databases: MySQL, PostgreSQL, MongoDB, Snowflake, Query Optimization, Data Warehousing
Analytics & BI: Tableau, Power BI, Statistical Analysis, Data Visualization, Business Intelligence

RELATED EXPERIENCE

Data Engineer Intern - Shreeji Woodcraft | Mumbai, India   **Jan 2024 - Jun 2024**
- Designed and built scalable ETL pipelines processing 10,000+ monthly records using Python and SQL, implementing comprehensive data validation frameworks and quality monitoring systems
- Developed dimensional data models and optimized database performance achieving improvement in query execution time, supporting analytics workflows for cross-functional business teams
- Collaborated with stakeholders to translate business requirements into technical data solutions, creating automated workflows that streamlined data processing and reporting operations
- Implemented data governance practices including documentation, lineage tracking, and metadata management, establishing foundation for reliable data-driven decision making across key business metrics

Web Development Intern - Vartak Computers Pvt Ltd | Mumbai, India   **May 2021 - Oct 2021**
- Developed REST APIs in Python/Flask handling 50,000+ monthly requests, implementing authentication, validation, and error handling for production web applications
- Built responsive web interfaces using JavaScript, HTML/CSS integrated with MySQL databases, delivering real-time data visualization and reporting capabilities
- Optimized database connections and SQL query performance improving application response times through efficient backend architecture and indexing strategies
- Implemented automated deployment workflows and version control practices, ensuring reliable code delivery and system maintenance for client-facing applications

PROJECTS

Movie Analytics Data Pipeline – Databricks, Alteryx, ER Studio, Power BI   **Jan 2026 - Feb 2026**
- Building end-to-end data pipeline processing 7 IMDb datasets with 10M+ records using Medallion architecture in Databricks
- Implementing data profiling and quality validation in Alteryx, tracking lineage from source to target systems
- Designing dimensional data model in ER Studio with fact and dimension tables for movie analytics and crew analysis
- Developing ETL workflows to unify multi-language titles, regional releases, and crew relationships across disparate sources
- Creating Power BI dashboards for movie trends, ratings analysis, genre comparisons, regional insights, episode metrics, etc.

F1 Championship Bayesian Predictor – Python, PyMC, Statistical Modeling   **Nov 2025 - Dec 2025**
- Built Bayesian hierarchical model using PyMC to predict 2025 F1 Championship standings for 20 drivers and 10 constructors
- Processed season data from 23 races including qualifying, sprint results, and weather conditions, establishing prior distributions updated with real-time data
- Executed Monte Carlo simulations generating probabilistic predictions with 72% accuracy for podium finishes

AWS Cloud Infrastructure Setup – AWS, Terraform, Networking   **Jan 2025 - Apr 2025**
- Deployed CI/CD pipeline with GitHub Actions, Packer, and Terraform reducing manual deployment effort by 60% and ensuring reliable, production-ready cloud deployments
- Provisioned secure AWS infrastructure across 3 availability zones and configured CloudWatch Agent in AMIs for monitoring 15+ metrics
- Enabled auto-scaling with Launch Templates and Auto Scaling Groups using CPU utilization-based policies
- Deployed Application Load Balancer to evenly distribute HTTP traffic across 4 EC2 instances and ensure 99.9% application fault tolerance
- Configured Route 53 with domain and subdomains (dev/demo) for DNS-based access and secure, load balanced routing