Hritvik Semwal

Boston, MA | (617) 516-4735 | semwal.h@northeastern.edu | LinkedIn | GitHub

EDUCATION

Northeastern University, Khoury College of Computer Sciences, Boston, MA
Master of Science in Data Science | 3.89 GPA *Jan. 2025 – May 2027*
Related Courses: Data Processing, Supervised Machine Learning, Deep Learning, LLMs

Vellore Institute of Technology (VIT), Bhopal, India
Bachelor of Technology in Computer Science Engineering | 3.8 GPA *May 2018 – May 2022*

TECHNICAL SKILLS

Languages and Libraries:	Python, SQL, R, Pandas, NumPy, Scikit-learn, SciPy, MLflow, PyTorch
Tools and Platforms:	Power BI, Tableau, Databricks, Spark, Azure, Airflow, Git, Docker, Excel
Focus Areas:	Content Analytics, Statistical Modeling, Hypothesis Testing, Data Visualization

PROFESSIONAL EXPERIENCE

Bajaj Finserv, Pune (India's leading NBFC serving 100M+ customers)

Data Scientist *Sep. 2023 – Dec. 2024*

- Built fuzzy decision tree models to route auto-rejected loan applications to alternative products, recovering $1.2M in previously lost conversions within 6 months by implementing product-specific evaluation logic.
- Automated manual JSON processing (4-6 hours) with Python/SQL ETL pipeline on Databricks featuring data quality validation and pre-calculated summary tables, reducing reporting time to minutes and computing costs by 20%.
- Created Power BI dashboards that identified growth opportunities using the open banking data integration system, referenced by C-level executives and cross-functional business teams for strategic decision-making.

Associate Data Scientist *Dec.2022 – Sep.2023*

- Enriched 500K+ customer records with mobile app, transaction, and credit bureau data (CIBIL, Equifax, Experian) using Python on Databricks, increasing approval rates by 27% through improved risk assessment.
- Segmented high-risk customers using k-means clustering and statistical models to refine credit policies, enabling $242K in additional safe lending opportunities previously excluded under conservative risk thresholds.

Data Technology Trainee (Intern) *Jan. 2022 – Dec.2022*

- Optimized inefficient marketing spends across communication channels by building Markov chain and ensemble models to identify high-converting touchpoints, improving retention rates by 14%.
- Automated call transcript NLP with spaCy and implemented MCP time series forecasting to produce summaries and optimal contact windows, removing manual logging and increasing call conversions by 20%.

ACADEMIC PROJECTS

Credit Risk Analysis | *Python, XGBoost, Decision Tree, SMOTE, Power BI, Streamlit*

- Built credit risk classification model using Lending Club data (2007-2018) to predict loan defaults, achieving 89% accuracy with Decision Tree after comparing 7 ML algorithms including Neural Networks.
- Addressed class imbalance using SMOTE and implemented feature engineering to reduce noise from 150+ features to top 5 predictors, improving high-risk recall from 68% to 76% while maintaining model interpretability for compliance.
- Deployed interactive Streamlit application with real-time risk scoring and Power BI dashboards for data quality tracking.

Subject-Invariant EEG Decoding | *Python, PyTorch, CNNs, Signal Processing*

- Developed deep learning solution for NeurIPS EEG Challenge addressing inter-subject variability in brain-computer interfaces, processing 80GB of multi-channel EEG signals (128 electrodes, 100Hz) for contrast change detection tasks.
- Designed hybrid architecture combining spatial Conv2D filters with Mixture of Experts (MoE) to capture cross-channel brain synchronization patterns, reducing baseline RMSE from 0.391 to 0.381 through systematic experimentation.
- Implemented leave-subject-out validation preventing subject-specific overfitting and optimized temporal-spatial feature extraction, enabling BCI models to generalize across diverse populations without lengthy calibration sessions.