Tommy Tang

25 Grenville Street, Toronto ON
(647)-607-2672 tommytang111@hotmail.com linkedin.com/in/zetang github.com/Tommytang111

SKILLS

Tools: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Tensorflow, Keras, PyTorch, OpenCV, NLTK, Gensim

Database Systems and Visualization: MySQL, PostgreSQL, MongoDB, Hadoop, Spark, Tableau, Alteryx, Azure

Languages & Technologies: Python, Java, Bash, HTML, CSS, Javascript, R, MATLAB

Machine Learning: Decision Trees, PCA, Clustering, NLP, CV, ANNs, RNNs, CNNs, LLMs, Transformers, GenAI

EDUCATION

Graduate Certificate in Applied AI Solutions – **George Brown College** *2024 – Present*

Master of Science in Molecular Genetics – **University of Toronto** *2024*

Bachelor of Science in Neuroscience – **University of Toronto** *2021*
Honours Degree with Physics Minor and Neuroscience Specialization | cGPA: 3.72/4.00

PROJECTS

Multi-Modal Stock Prediction Analysis| *Python*, *TensorFlow, Transformers, RNNs, LSTM, CNN, LLMs* *01/25-02/25*

- Designed and implemented an end-to-end system that combines deep learning models and news sentiment analysis techniques to improve the interpretability and accuracy of stock price movement predictions.
- Compared the performance of Vanilla RNN, Encoder-decoder RNN, LSTM (RNN), and CNN for capturing stock temporal trends and fine-tuned baseline BERT to an accuracy of 0.983 for financial news sentiment analysis.
- Demonstrated that integrating advanced sentiment analysis with traditional time-series forecasting methods can provide actionable insights by correlating sentiment shifts with market behaviour.

Hybrid Movie Recommendation System | *Python, NumPy, Pandas, Matplotlib, Scikit-Learn, SVD, NLP* *11/24 – 12/24*

- Developed a hybrid movie recommendation system combining content-based filtering (with TF-IDF and cosine similarity) and collaborative filtering (with SVD) to provide personalized movie recommendations.
- Implemented data preprocessing and feature extraction techniques on a real dataset of 20M user ratings.
- Evaluated model performance achieving a 0.749 MAE for ratings predictions and visualized cumulative variance to determine optimal latent dimensions.

Prescription Information Extraction | *Jupyter, Python, NLTK, CRF, NER, NLP* *10/24 – 11/24*

- Built a Named Entity Recognition (NER) model using Conditional Random Fields (CRF) to extract prescription information from medical text, improving data extraction accuracy.
- Designed and implemented feature extraction techniques, including POS tagging with NLTK and custom token features, to enhance model performance.
- Trained and evaluated the CRF model on labeled prescription data, achieving high precision and recall in identifying key prescription components.

Credit Card Approval Prediction | *Jupyter, Scikit-Learn, Matplotlib, Pandas, NumPy, GridSearchCV* *09/24 – 10/24*

- Developed and optimized classification models (Random Forest, Logistic Regression, Decision Tree, SVM, and SGD) to predict credit card approval outcomes for clients based on personal and financial data.
- Conducted extensive data preprocessing, including data imputation, one-hot encoding, and feature scaling.
- Utilized GridSearchCV for hyperparameter tuning and SMOTE for overcoming class imbalances.
- Concluded that Random Forest was the best model for credit approval classification, having optimized model performance to an ROC-AUC score of 0.72.

Infarct Prediction Using CT Perfusion | *R, ggplot2, randomForest, Logistic Regression, ROC-AUC* *09/20 – 04/21*

- Utilized logistic regression and receiver operating characteristic (ROC) curve analysis to evaluate computed tomography perfusion (CTP) parameters for predicting ischemic core infarcts in stroke patients.
- Discovered optimal thresholds for cerebral blood volume decreased from 1.15 to 0.35ml/100g after 3 hours of stroke, providing radiologists with tools for more precise ischemia assessments.
- Concluded that the accuracy of prediction improved with greater stroke onset to imaging times in real medical datasets, highlighting the importance of this metric for thrombectomy treatments.

WORK EXPERIENCE

Research Associate | Mount Sinai Hospital **|** Toronto, ON *05/24 – 08/24*

Investigated the localization of the sodium leak channel subunit FAM155A and other NALCN-related pathophysiology

- Utilized Seurat pipeline to perform clustering with uniform manifold approximation and projection (UMAP) on public scRNAseq datasets and found co-localization of NALCN subunits in Purkinje cells of the cerebellum.
- Mentored 3 new graduate students, teaching them experimental, analytical and scientific presentation skills.

Research Internship | SickKids Hospital **|** Toronto, ON *05/21 – 08/21*

Investigated the impact of pediatric brain tumours on structural connectivity and correlations with cognitive function.

- Created a custom MRTrix Diffusion Tensor Imaging pipeline using Bash and Python to process white matter metrics across 300000+ brain images spanning 5TB of cloud data.
- Utilized R and MATLAB to perform PCA and statistical significance testing, reporting that pediatric brain tumor patients had significantly compromised integrity of white matter tracts compared to healthy controls.