Camellia Debnath

(857)930-3409 • debnath.c@husky.neu.edu • www.linkedin.com/in/camellia-debnath • github.com/dn-cam • https://paranoidandroid-com.webnode.com/

EDUCATION

Northeastern University, Boston, USA *Sep'18 – Aug'20*
Khoury College of Computer and Information Sciences
Master of Science in Data Science

- **GPA**: 3.92/4.00
- **Relevant Courses**: Algorithms (**Teaching Assistant** for Spring'19, Spring'20), Data Management and Processing, Supervised Machine Learning, Foundations of Artificial Intelligence, Data Mining and Unsupervised Machine Learning.

BITS Pilani, Pilani, India
B.E. (Hons.) Electrical and Electronics *Aug'11- May'15*

TECHNICAL KNOWLEDGE

Languages: Python, R, Java, C, PL/SQL, Angular 2, Typescript, Javascript
Concepts: Regression Modeling, Classification, Clustering, Language Modelling, Data Wrangling, Exploratory Data Analysis, Model Selection & Assessment, Topic Modelling, Computer Vision, Agile, Unix/Linux, Relational Databases
Tools: Eclipse, UDeploy, RStudio, Jupyter Notebook, Autosys, Git, Jira
Libraries numpy, pandas, scikit-learn, seaborn, TensorFlow, NLTK, tidyverse, dplyr, ggplot2, OpenCV

WORK EXPERIENCE

Software Developer, CITICORP, Pune, India *Jul'15 – Jul'18*

Natural Language Processing Framework for Automated Generation of Test Cases

- Designed a framework using **Apache Jena** and **Stanford NLP** libraries for parsing plaintext, performed POS-Tagging.
- Created **RDF** knowledge graphs, traversed the knowledge graph to generate **gherkin** scripts.
- Fed gherkin scripts into Cucumber Test runner, generated test cases which were then run on the application code.
- This reduced communication latency, and increased average burndown efficiency factor by 35%

Real-Time Processing of Big Data to process Audit Data

- Designed a **Java** framework using **Apache Storm** topologies and **Apache Kafka** topics for processing trade exceptions.
- Saved processed errors into **MongoDB** for reference, and possible resolution.
- Transformed +1 day batch processing to real-time error resolution, garnering customer satisfaction;

Predictive Model for Brokerage Fee Calculation

- Analyzed 10+ years of brokerage fee reports using **dbplyr**, **dplyr, ggplot2** libraries, performed **Exploratory Data Analysis.**
- Developed a **linear regression** model to predict brokerage fees, and inform business user before executing a trade.

GRADUATE RESEARCH FELLOWSHIP

Northeastern University *Sept'19 – Present*

Automated Problem Solving Using Neural Monte Carlo Tree Search (Ongoing Individual Research)

- Used Google DeepMind's **AlphaZero** algorithm for solving combinatorial optimization problems, eg the Highest Safe Rung.
- Found relevant **hyperparameters** (number of self-plays, exploration rate, MCTS sims) and their effect on the learning rate.
- Formulated game switch statistics that resulted into an important discovery about the effect of TensorFlow version difference.
- Generated game-trees to explore how the winning strategy evolves for both proponent and opponent in the game
- Created HSR (Highest Safe Rung) environment in OpenAI Gym for evaluation of Neural MCTS algorithm.

ACADEMIC PROJECTS

Quora Insincere Question Classification *May'19 – Aug'19*

- Analyzed Quora question text dataset to detect insincere content using binary classification.
- Performed **TF-IDF vectorization**, **Sentiment Analysis** using Python **NLTK** framework for gauging overall sentiment.
- Compared performances of **SVM** (Support Vector Machines), **CNN** (Convolutional Neural Networks) and **LSTM RNN** (Long Short-Term Memory Recurrent Neural Network) for classifying.

Bankruptcy Prediction Using Various Classifiers *Jan'19 – Apr'19*

- Analyzed 5 years of Polish Companies Dataset, containing 64 econometric ratios and bankruptcy labels.
- Handled missing data using **Mean value imputation**, and **SMOTE** (Synthetic Minority Oversampling Technique) on training data for handling imbalance between two classes (bankrupt and non-bankrupt).
- Calculated **Correlation matrix**, performed **Cross-validation** for feature sub-setting and hyperparameter tuning.
- Fit Logistic Regression, Naïve Bayes, **LDA** (Linear Discriminant Analysis), **QDA** (Quadratic Discriminant Analysis), **SVM** (Support Vector Machine) and feed forward Neural Networks to compare various classification techniques.

Image Building and Recognition using Unsupervised Techniques on CIFAR and 20NewsGroups dataset *Sept'18 – Dec'18*

- Calculated **pairwise similarity matrices** using cosine, Euclidean, Jaccard, Editing and Manhattan similarity.
- Applied **feature selection** techniques – **PCA**, **ChiSquare**, **Mutual Information**, **L1** and **HARR** feature extraction.
- Evaluated **clustering** performance - **KMeans**, **Gaussian Mixture Models**, **DBSCAN** and **Hierarchical** Clustering on datasets using metrics like **Purity** and **Gini index**.
- Re-performed the clustering tasks with **TensorFlow** as an **AutoEncoder**, indexed both datasets using **Elastic Search** and used it for **Topic Modelling** – **LDA** (Latent Dirichlet Allocation**)** and **NMF** methods.