SHARVIL SALIL PRADHAN

Email: spradha5@ncsu.edu | Phone: 919-931-8183 | LinkedIn: linkedin.com/in/sharvil-pradhan | Website: sharvilpradhan.github.io

EDUCATION

North Carolina State University | Raleigh, NC, USA — May 2020
Degree: Master of Computer Science — Major: Computer Science — GPA: 3.83/4
Coursework: Neural Networks & Deep Learning | Automated Learning & Data Analysis | Artificial Intelligence | Natural Language Processing | Algorithms & Data Structures | Experimental Statistics | Database Management and Systems | Human-Computer Interaction | Bioinformatics

Sardar Patel Institute of Technology (affiliated to University of Mumbai) | Mumbai, India — May 2018
Degree: Bachelor of Engineering — Major: Electronics Engineering — GPA: 9.18/10
Coursework: Object-oriented Programming | Structured Programming

TECHNICAL SKILLS

- **Programming Languages:** Python (NumPy, Pandas, Scikit-learn, SciPy), Java, Julia, C, C++
- **NLP Tools:** NLTK, Gensim, SpaCy, TextBlob
- **Statistical Tools:** R, JMP (SAS)
- **Machine Learning Frameworks**: TensorFlow, Keras
- **Data Visualization:** Tableau, Matplotlib, Seaborn, ggplot
- **Database Querying Language:** SQL
- **Database:** Oracle, MySQL
- **Operating Systems**: Linux, Windows
- **Tools:** Jupyter, RStudio, Apache Spark, GitHub, Microsoft Excel

ACADEMIC PROJECTS

Data Science Projects:

- **Predictive Modeling – Estimating Persuasiveness**
  - Designed a system to award points to logical comments on social media thereby filtering out spam messages
  - Extracted unstructured text comprising 5000 Reddit 'Change My View' Discussion threads and pre-processed the noisy data, including clean and prune data, handle missing values, under-sample biased data
  - Conducted feature engineering to extract important features (politeness, sentiment, hedge word count) from raw text for awarding points to logical comments
  - Enhanced accuracy from 54% to 77% by implementing deep learning model over classical machine learning algorithms (Support Vector Machines and Decision Trees)
  - This system saves the users' time by displaying comments in decreasing order of relevance to the topic
  - Estimating Persuasiveness of Online Comments - Source Code
- **Data Visualization/Analysis – Liver Disease Forecasting**
  - Generated model directed towards predicting possibility of a liver disease in a person
  - Detected outliers by generating boxplots and analyzed trends in data through Tableau visualizations
  - Developed linear regression models to determine factors contributing most towards occurrence of liver disease and tested hypotheses to check working of different models
- **Machine Learning – Event Timeline Detection**
  - Objective was to manufacture a model that understood cause and effect relation between two sentences
  - Quantified performance of SVM and Random Forest supervised learning algorithms in detecting causality between event pairs for NLP word embedding techniques such as word2vec, SpaCy and GloVe
  - Achieved 30% improvement in precision of cause and effect detection between event pairs by executing modification of doc2vec word embedding technique using Gensim
  - Event Timeline Detection - Source Code

- **Data Analytics – Website Redesign**
  - Collaborated with UX developers to work on redesigning the Computer Science website of NC State University
  - Analyzed feedback surveys to draw insight into user behavior, identifying new opportunities for process improvement
  - Displayed clear data visualizations in the form of charts and graphs for conveying comparison of multiple user stories and their needs
- **Image Classification – Leaf Wilting Identification**
  - Aimed at discovering early signs of disease in soybean plants based on the appearance of their leaves
  - Built a Convolutional Neural Network to detect signs of wilting and extent of disease in soybean leaves based on soybean plant images
  - Improved accuracy of classification from 41% to 82% on utilizing semi-supervised learning approach along with transfer learning through VGG-16 model
  - Leaf Wilting Identification - Source Code
- **Research – Adversarial Attack on Sentiment Analysis models**
  - Tested our hypothesis that the word embedding models do not have enough representations of low frequency words
  - Implemented POS tagging to identify adjectives and verbs from Zomato food reviews and IMDB movie reviews
  - Discovered target words via VADER sentiment analysis tool and selected synonym having minimum cosine similarity for the word embedding using word2vec
  - Altered the input by replacing target words in the reviews with their synonyms that were farthest to them in NLP vector representation
  - Examined effect of altered reviews on the accuracy and confidence of sentiment classification state-of-the-art LSTM model
  - Adversarial Attack on Sentiment Analysis models - Source Code
- **Language Modeling – Caption Generation for Images**
  - Built an LSTM model to generate textual captions for images in the Flickr8k dataset
  - Evaluated the generated captions using BLEU scores for unigram, bigram, trigram and 4-gram language models
  - Tested efficacy of the model for text-based image search
  - Attained multiple tasks using the same neural network model, hence saving computing memory and resources
  - Caption Generation for Images - Source Code

Database Management Project:

- **RDBMS – Automobile Service Center Management**
  - Purpose of this project was to create and manage a database system for a car service center to store customer information and perform car maintenance
  - Architected a relational database model (E-R diagram) for managing a car service center database
  - Engineered Java application flow utilizing JDBC to communicate with Oracle database
  - Achieved functionality for tasks by writing optimal SQL queries, procedures, and triggers
  - Car Service Center Database Management - Source Code

Software Development Project:

- **Multi-objective Optimization – Motif Discovery in DNA**
  - Devised software to find motifs (patterns) in sequences of DNA
  - Obtained 98% accuracy to find implanted motifs in a genome sequence by deploying multi-objective optimization Artificial Bee Colony (ABC) algorithm and Gravitational Search Algorithm (GSA)
  - Revamped ABC algorithm to include negative feedback, thus enhancing its operation for motif lengths above 20
  - Assessed behavior of ABC, GSA and modified ABC algorithms on artificial genome sequences and real datasets (TRANSFAC database)
  - Motif Discovery in DNA - Source Code

RESEARCH PUBLICATION

Pradhan, S. “Cloud-based Smart Parking System". International Conference on Inventive Communication and Computational Technologies (ICICCT) 2018.
Published in IEEE Xplore Journal