Suhas Maddali

1203 Boylston Street,
Boston, MA 02215
Email: suhas.maddali.edu@gmail.com
Phone: 4802863087
LinkedIn: https://www.linkedin.com/in/suhas-maddali-b9b146136/
GitHub: https://github.com/suhasmaddali
Kaggle: https://www.kaggle.com/suhasmaddali007
Medium: Suhas Maddali – Medium

Objective

A Passionate Data Science and Machine Learning Graduate looking forward for a role as a Data Scientist Intern in planning, production, analysis, optimization, visualization, pre-processing, deploying machine learning and deep learning models to produce the optimum results and fine tuning the models to get the best accuracy.

SKILL SET	PROFILE SNAPSHOT
Python R Natural Language Processing (NLP) Git Probability and Statistics Machine Learning Optimization Data Visualization Feature Engineering Data Manipulation SQL Deep Learning Typing Expert (100 WPM)	• Worked with Machine Learning and Deep Learning Algorithms over 2 years and created a list of projects in my blog that includes Real-World Case Studies. • Experience coding in Python, R and SQL needed in the field of Data Science and Analytics. • Strong theoretical and practical knowledge of Machine Learning and Deep Learning techniques used in the process of Prediction and Analytics. • Successfully deployed Machine Learning and Deep Learning Models to get the best accuracy in the predictions of the models. • Worked with data sets that required Natural Language Processing (NLP) and extracted insights from texts. • Used Data Cleaning and Manipulation methods to generate meaningful insights from data. • Strong theoretical knowledge of Probability and Statistics used in generating key insights from data. • Strong at various Visualization Libraries in Python such as Matplotlib, Seaborn and Plotly. • Excellent Written and Verbal Communication Skills needed in the Business Environment. • An Expert Typist with a speed averaging about 100 Words Per Minute (WPM).

Technical Skills

Machine Learning Algorithms:	Support Vector Machines (SVM), Linear Regression, Logistic Regression, Naïve Bayes, k-Nearest Neighbors (KNN), Decision Trees, Random Forests, Bagging, Boosting, Xgboost, Gradient Boosting Decision Trees (GBDT), DBSCAN, K Means Clustering, K++ Clustering, Agglomerative Clustering
Deep Leaning Methods:	Convoluted Neural Networks (CNN), Recurrent Neural Networks (RNN), Auto - Encoders, Long Short-Term Memory (LSTM)
Optimization Algorithms	Adagrad, Adadelta, Adam, RMS Prop, Nesterov Accelerated Gradient (NAG)
Dimensionality Reduction:	Principal Component Analysis (PCA), t – SNE, Truncated SVD
Natural Language Processing:	Bag-of-Words, TFIDF Vectorizer, Word2Vec, TFIDF Word2Vec, BERT, RoBERTa
Utilities:	Scikit-Learn, NumPy, Keras, TensorFlow, Seaborn, Matplotlib, Glove Vectors, Vectorizers, Matplotlib, Scipy, Spacy, NLTK, WordCloud
Projects Completed:	Washington Bike Demand Prediction, Car Prices Prediction, Twitter Sentiment Analysis, YouTube Video Analysis, Heart Disease Prediction, MNIST Using Keras, Steel Defect Detection, Reddit Comments Analysis, Cancer Diagnosis Prediction, Home Credit Loan Sanction Prediction, Human Activity Recognition

Portfolio/Projects Completed

Washington Bike Demand Prediction

Objective: The objective of this project was to predict the demand for bikes during different instances of time by taking into consideration the number of casual riders, the season, weather and other factors that impact the demand respectively.

Tools: Jupyter notebook

Duties: Performed Exploratory Data Analysis (EDA) to understand the demand for bikes during different seasons and months. Furthermore, the distribution of some features such as windspeed, temperature and humidity are also plotted to get an understanding of the spread of data. Scatterplots were also used to see the correlation between different features such as registered users and casual riders. In addition to this, various machine learning algorithms were used such as Deep Neural Networks, K Neighbours Regressor, PLS Regressor, Decision Tree Regressor, Gradient Boosting Regressor and Logistic Regression.
Finally, a Scatterplot between the actual values and the predicted values on the cross-validation data was plotted to understand the performance of different machine learning models.

Outcome: The models that performed the best without hyperparameter tuning were Deep Neural Networks and Gradient Boosted Decision Trees. Based on the results, Deep Neural Network was taken as the main model, and performing Hyperparameter Tuning led to a good increase in the performance with the Mean Absolute Error reducing from 130.76 to 22.7 respectively.

Result: The Deep Neural Network model was able to do exceptionally well on the test set with the Mean Absolute Error about 25 and the Mean Squared Error of about 1800 respectively. Below is the link to the notebook.
[suhasmaddali/Washington-Bike-Demand-Prediction: Using different machine learning and deep learning techniques, the demand for Washington Rental Bikes would be predicted so that the companies could ensure that people receive bikes as and when needed. (github.com)](#)

Heart Disease Prediction Using Machine Learning and Deep Learning

Objective: The main objective of the project was to predict the chances of a heart disease in a patient using machine learning and deep learning algorithms given some of the important metrics such as cholesterol levels, blood pressure and so on.

Tools: Jupyter notebook

Duties: As part of a team of 4 members, my role was to guide my teammates in driving the project forward. I had to first pre-process the data that would later be used in the machine learning algorithm. I was the team lead and I divided the work between my teammates. I learned machine learning theory before implementation of the code and did hyperparameter tuning to get the optimum results.

Outcome: The model was able to predict the chances of a heart disease in a patient with an accuracy of about 98 percent. In addition to this, other models such as Logistic regression and Shallow Neural Networks were used but results were not as close to that of the Deep Learning Model. In addition to this, this project was presented to our project guide and he told us that it was really good. We developed a project report highlighting the various Machine Learning and Deep Learning Algorithms that were used in the process of prediction. Data visualization, theory and the code were also documented in the report.

Result: Obtained high distinction and full credits for the project in the final year of engineering. In addition to this, I also got a bronze medal in Kaggle after uploading this code with the additional description. Below is a link to the notebook.
[suhasmaddali/Heart-Disease-Prediction: We would be using machine learning models to predict the chances of a patient suffering from a heart disease using various features such as cholesterol levels and chest pain type. We would be just considering a sample dataset just to get an understanding of the various machine learning models that could be put to action and learn their implementation. We would also be considering various classification metrics just to compare how well the models did on the test data. (github.com)](#)

Predicting Readability of Texts Using Machine Learning

Objective: There can be a lot of books and articles being published every minute. It would be really good if the difficulty of these texts is predicted with the aid of Machine Learning and Data Science. The objective of this project was to determine the Difficulty of Texts Using Machine Learning and Deep Learning.

Tools: Jupyter notebook

Duties: Exploratory Data Analysis was performed to get a good understanding of the data. Various plots such as Missingno plots, Histograms and WordCloud were used to get to know the total number of Null values and the Text Word frequency. Since the data that was available contained only Text Information along with Standard Error values, new features had to be created to improve the performance of the Machine Learning Models.

Various functions that create new features and perform Pre-processing were defined. Furthermore, Vectorization and Plotting functions were also used in the process.

There were Numerous Machine Learning Models used to predict the difficulty of Texts. The Machine Learning Models used were Neural Network Model, Linear Regression, K – Neighbours Regressor, PLS Regression, Decision Tree Regressor and Gradient Boosting Regressor.

In addition to using TFIDF and Count Vectorizes, Word2Vec and TFIDF Word2Vec Vectorizers for replacing the text with vectors were used. This generated a 300-Dimensional vector (Glove Vectors) containing the similarity matrix of different words which were later given for predictions.

Outcome: It could be seen from the results that Deep Neural Networks and Gradient Boosted Decision Trees performed the best for predicting the difficulty of different texts. One of the most interesting findings in the project is that when using vectorizers such as Word2Vec, the performance of the Machine Learning Models improved significantly. Generally, Word2Vec (Glove Vectors) preserve the Semantic Meaning of the text compared to the Count Vectorizers and TFIDF Vectorizers. The Validation loss of the Deep Neural Network with 10 epochs went from about 0.49 to 0.29 with transitioning from TFIDF vectorizer to Word2Vec respectively. This clearly shows how using Word2Vec in Natural Language Processing (NLP) tasks improves the performance of the Machine Learning and Deep Learning Models.

Result: There was a significant decrease in the Mean Squared Error and Mean Absolute Error when changing the features created by Count Vectorizer and TFIDF Vectorizer as compared to Word2Vec and TFIDF Word2Vec respectively. Below is a link to the project on GitHub.

suhasmaddali/Predicting-Readability-of-Texts-Using-Machine-Learning: The aim of this project is to predict the difficulty level of the text using various machine learning techniques respectively. (github.com)

Education

Masters in Data Science   **May 2021 – Present**
Northeastern University Khoury College of Computer Sciences
Boston, Massachusetts
Marks Scored: 91%

Bachelor of Technology   **May 2015 – April 2019**
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, India
Marks Scored: 90%

High School   **June 2013 – May 2015**
Narayana Junior College
Hyderabad, India
Marks Scored: 97%

Junior High School   **August 2009 – May 2013**
Vikas The Concept School
Hyderabad, India
Marks Scored: 90%

Internship/Work Experience

Machine Learning Intern   **April 2018 – December 2018**
Solbots Technologies Private Limited
Roles and Responsibilities:
- Collected the data needed for machine learning and deep learning algorithms.
- Used various deep learning algorithms and tested the accuracy with different hyperparameters.
- Trained the machine learning and deep learning models with many images for multi-class classification.

Certifications / Trainings

Certified:
- o Deep Learning Specialization in Coursera
- o Machine Learning professional Certificate by Stanford University
- o Complete Python Bootcamp from Udemy
- o Python for Data Science and Machine Learning Bootcamp from Udemy
- o HTML and CSS Certification
- o Applied AI Certification

Blogs

"Introduction to Natural Language Processing for Machine Learning" - [Introduction to Natural Language Processing for Machine Learning | by Suhas Maddali | MLearning.ai | Aug, 2021 | Medium](#)

"Different Types of Supervised Machine Learning Models" - [Different types of Supervised Machine Learning Models | by Suhas Maddali | Jul, 2021 | Medium](#)

"What are Convoluted Neural Networks (CNN)?" - [What are Convoluted Neural Networks (CNN)? | by Suhas Maddali | Nerd For Tech | Medium](#)

"Why is Hyperparameter Tuning Important in Machine Learning" - [Why is Hyperparameter Tuning important in Machine Learning? | by Suhas Maddali | Nerd For Tech | Aug, 2021 | Medium](#)

"Predicting the Sentiment of a Text with Machine Learning" - [Predicting the Sentiment of a Text with Machine Learning | by Suhas Maddali | MLearning.ai | Aug, 2021 | Medium](#)