Ever since I started learning and exploring about Machine learning, I have done many projects on various data sets and this document consists of brief summaries of them. I have also included certifications of my learnings and participations in this field.

Data science Projects

Chaitanya varma

Machine learning Projects

- COVID-19 analysis and prediction
  Number of Infected cases, recoveries and deceased are analyzed and SARIMAX model is built to forecast these variables. The stationarity of the variables is checked by plotting rolling mean and rolling standard deviation and also Dicky fuller test is used. ACF and PACF plots have been analysed to choose the degree of Auto regression (p), differencing(d) and moving average (q). We have considered (0,1,2) as possible values for each of p,d,q and P,D,Q. Model is built with all possible combinations (27*27) and corresponding AIC values are calculated. The combination with which the model's AIC is least is selected. This work has been presented in **Machine Learning as a Service to Industries (MLSI) 2020 National level E-conference by AICTE**
  Key skills: Data preprocessing, Time series analysis, parameter tuning.
  Report:
  https://github.com/chaitanya644/Puzzle-box/blob/master/Final%20Report.pdf

- Mutual fund rating
  Used ML techniques from python to predict the ratings of the mutual fund using data cleaning and merging multiple data sets that have mutual fund performance after 1 year, 3 year and 5 years. Highly correlated independent features were removed and Anova has been used to check the classification ability of the features. Missing values are appended with median and range. Decision Tree classifier, KNN and Random Forest are fit on the data and results are predicted. As Random forest has produced better f1-score for all the categories, it was chosen to predict the ratings of the unknown mutual funds.
  Key skills: ANOVA, KNN, Decision Trees, Random Forest.

- Book Recommendation
  Titles of the books are vectorized and cosine similarity between them is calculated. Top 10 highly similar titles of books are recommended to the user based on titles of previously read books.
  Key skills: Stemming & Lemmatization, Tfidf vectorizer, SVD.

- Loan Defaulter Prediction
  Important features that have impact are identified and ML Classification model is built to predict loan defaulters for the loan accounts which yet to be closed.
  Key skills: Decision Tree, Regularization, Gradient Boost, Random Forest.

- Play store games price prediction
  Based on features like primary and secondary genres, size of the game, developer and user ratings and release dates game prices are predicted.DecisionTree Regressor and Random forest were fit on the data.
  Key skills: Encoding of Categorical variables,Decsion Tree, Random Forest.

Natural Language Processing projects

- Topic Modelling of Quora questions
  Tfidf vectoriser is used to vectorise the questions and stop words are removed in the process. Both NMF(Non negative matrix factorization) and LDA(Latent Dirichlet Allocation) were used seperately to categorise the questions and top 15 words of a topic are used to describe the title of the topic
  Key skills: Tfidf vectoriser, LDA, NMF.

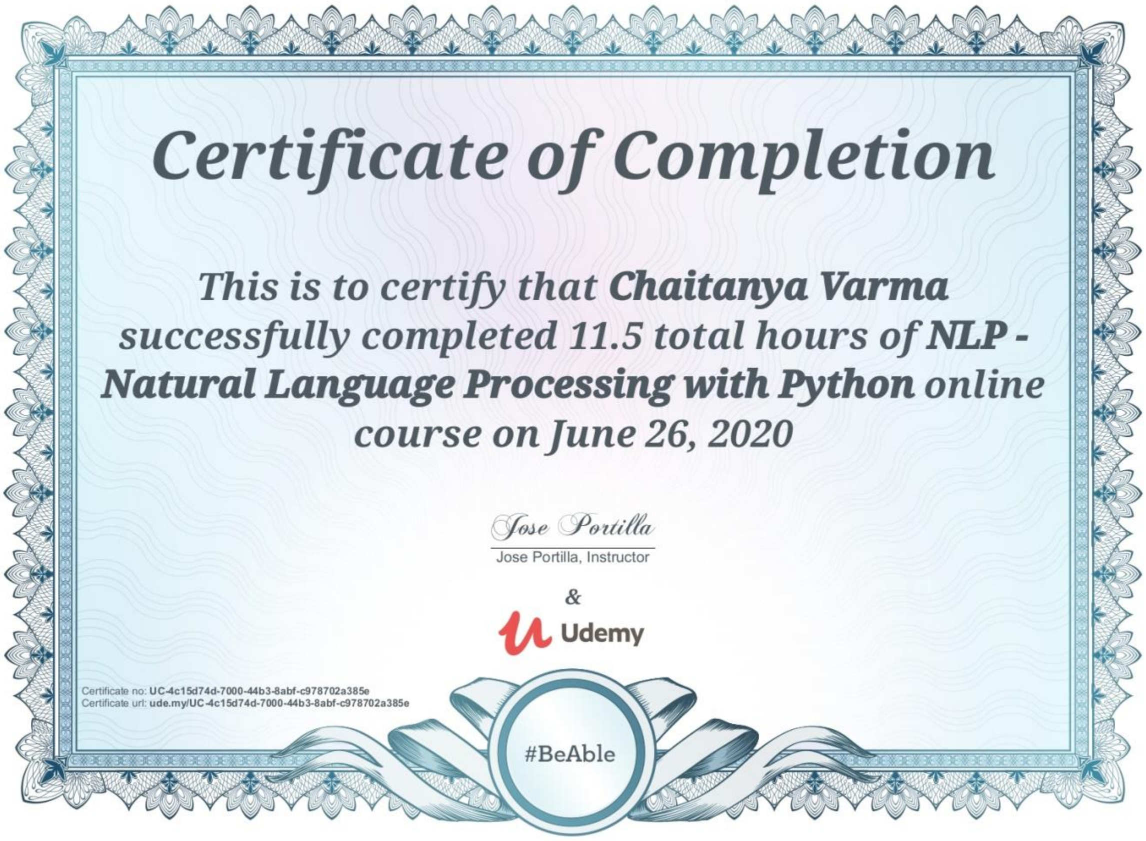
- Text generation using LSTM
  Moby dick chapters text is used and words are tokenized. First 50 words are used to predict the 51st word. Sequential model is created with two layers of LSTMS between dense layer and embedding layers. Each individual word is treated as a category for the model to predict the next word.
  Key skills: Keras Tokenisation, Neural network for text generation.

- Implementation of “End to End Memory networks” publication by Facebook AI research
  Babi data set from Facebook research is taken. Each record contains story, question and answer. Vocabulary is built and tokenized. All stories, questions and answers are vectorised. Single layer version of publication described model is built and answers were predicted for given story and question.
  Key skills: Tokenisation, vectorisation, Connecting Multiple sequential neural networks.

- Sentiment analysis of movie reviews
  Sentiment intensity analyser from nltk sentiment Vader has been used to get polarity scores of the reviews and compound value is considered to determine sentiment of the review. Compound value of reviews which are closer to +1 is classified as positive and those which are closer to -1 are classified as negative.
  Key skills: python iterators(intertuples), Sentiment intensity analyser.

- Text classification using SVC
  Same movie review dataset used in sentiment analysis is taken and reviews that are already labelled as positive and negative are fit into Linear SVC model and predictions are made on the test set. It is seen that SVC has classified the reviews with higher accuracy than Vader sentiment analyser.
  Key skills: Pipelines in python, Tfidf vectoriser, Linear SVC.

NLP Certification



Certificate of Completion

This is to certify that ***Chaitanya Varma*** *successfully completed 11.5 total hours of* ***NLP - Natural Language Processing with Python*** *online course on June 26, 2020*

Jose Portilla

Jose Portilla, Instructor

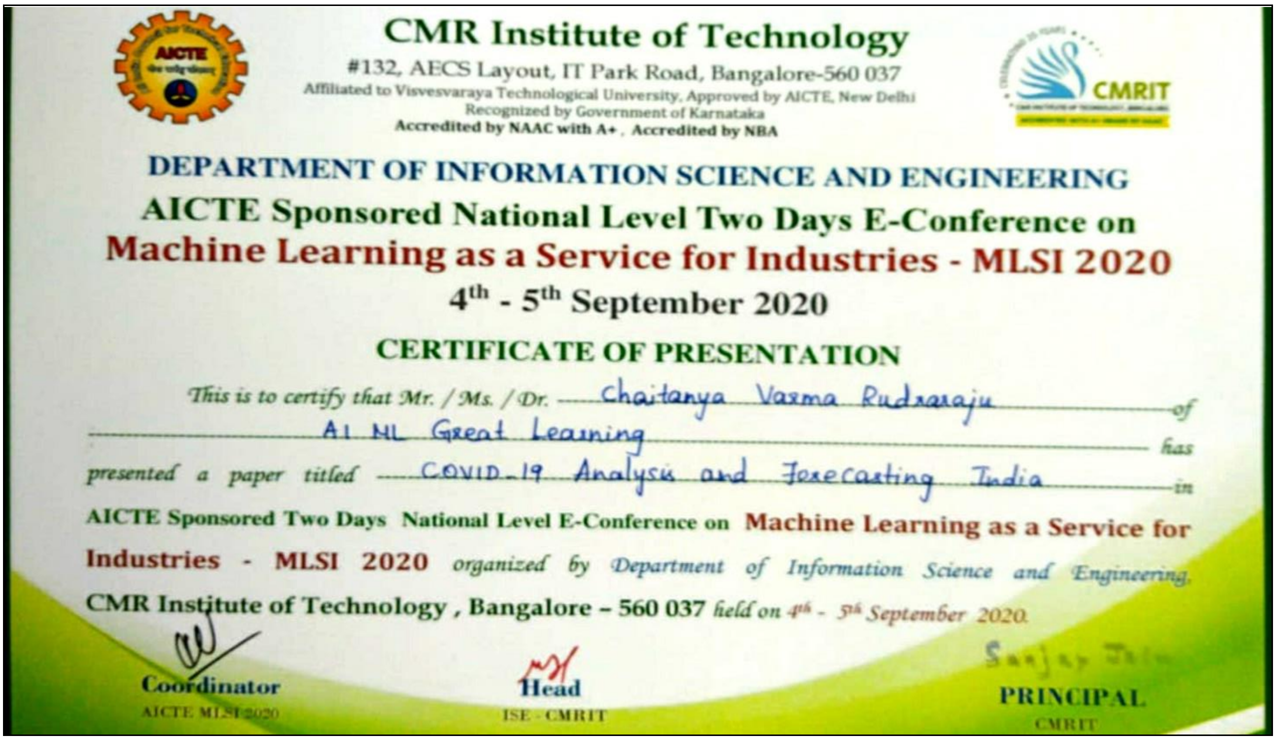
&

Udemy

Certificate no: **UC-4c15d74d-7000-44b3-8abf-c978702a385e**
Certificate url: **ude.my/UC-4c15d74d-7000-44b3-8abf-c978702a385e**

#BeAble

Participation in National E-Conference



AICTE

CMR Institute of Technology

#132, AECS Layout, IT Park Road, Bangalore-560 037
Affiliated to Visvesvaraya Technological University, Approved by AICTE, New Delhi
Recognized by Government of Karnataka
Accredited by NAAC with A+ , Accredited by NBA

CMRIT

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

AICTE Sponsored National Level Two Days E-Conference on

Machine Learning as a Service for Industries - MLSI 2020

4th - 5th September 2020

CERTIFICATE OF PRESENTATION

This is to certify that Mr. / Ms. / Dr. Chaitanya Varma Rudraraju *of* AI ML Great Learning *has presented a paper titled* COVID_19 Analysis and Forecasting India *in* **AICTE Sponsored Two Days National Level E-Conference on Machine Learning as a Service for Industries - MLSI 2020** *organized by Department of Information Science and Engineering,* **CMR Institute of Technology , Bangalore – 560 037** *held on 4th - 5th September 2020.*

Coordinator
AICTE MLSI 2020

Head
ISE - CMRIT

Sanjay Jain
PRINCIPAL
CMRIT