Mounika Marreddy

Curriculum Vitae

IIITH, LTRC Lab
Hyderabad, 500032
(+91) 9553127911
mounika.marreddy@research.iiit.ac.in

Areas of Interest

Machine Learning.

Natural Language Processing.

Education

2018–Now **PhD**, *IIITH*, *CGPA - 8.33/10* .

2014–2016 **M.Tech in Artificial Intelligence**, *University Of Hyderabad*, *CGPA - 7.8/10* .

2010–2014 **B.Tech in CSE**, *Nalanda Institue Of Engineering & technology*, Guntur, *Percentage - 77.86*.

Publications

OCT 2020 – **Clickbait Detection in Telugu! Building from Scratch?**, *(EACL-2021), (CORE RANK: A)*, In Review.

MAY 2020 – **Am I a Resource-Poor Language? Datasets, Embeddings, Models and Analysis for four different NLP tasks in Telugu Language**, *(Computational Linguistics Journal), (CORE RANK: A*)*, In Review.

OCT 2020 – **Multi-Task Text Classification using Graph Convolutional Neural Networks for Resource-Poor Language**, *(NIPS-2020,WiML workshop), (CORE RANK: A*)*, Accepted.

OCT 2020 – **Unsupervised Graph based Telugu News Articles Text Summarization**, *(NIPS-2020,WiML workshop), (CORE RANK: A*)*, Accepted.

AUG 2019 – **Evaluating the Combination of Word Embeddings with Mixture of Experts and Cascading gcForest in Identifying Sentiment Polarity**, *(KDD WISDOM workshop -2019), (CORE RANK: A*)*, Accepted.

DEC 2018 – **Affect in Tweets using Experts Model**, *(PACLIC -2018), (CORE RANK: B)*, Accepted.

SEPT 2017 – **Multi-Arm Active Transfer Learning for Telugu Sentiment Analysis**, *(IAL Workshop @ ECML PKDD-2017), (CORE RANK: A)*, Accepted.

Teaching Assistant Work

August 2020 – Current **Teaching Assistant**, *Advanced Natural Language Processing*, IIT-Hyderabad, Mentoring 4 teams to apply Machine Learning techniques on problems related to Summarization, hate speech and Sentence simplification wiki in NLP area. Taught online tutorials, designed assignments and handled evaluation of assignments for this course..

Research Assistant Work

August 2020 – Current **Research Assistant**, IIITH.
I am part of the Quality Assessment team for the development of a parallel corpus for ILMT (Indian language to Indian language Machine Translation) project. Currently, for the Hindi-Telugu Machine Translation system, the corpus is being prepared. Data is collected from domains like Education, News, Health, Biology, and Law. My role is to check the Hindi data collected and translated into Telugu by different companies. Suggestions are commented on translated sentences by verifying an average of 1000 sentences per week.

Work Experience

July 2016 – December 2017

Data Scientist, QUADRATYX, Hyderabad.

I worked on Route Optimization project for middle east client. Here the goal of the project is to recommend optimized route for the sales person who delivers goods for each store. My responsibilities in this project include, data preprocessing, creating unified view from all data sources, data visualizations using Tableau and model building. Worked on data preprocessing techniques for identifying features from Scanned documents, Search-able PDF's, Images and News Papers and used these features for model building. Worked on Product order recommendation project, which includes Data Visualization, Feature Selection and Model building.

Industry Projects

Route Optimization for salesman

Python, SQL, Tableau

Route Optimization is used to predict the optimized route for the sales person who delivers goods for each store based on the information like store details, warehouse details and some other constraints provided by the client. Here we constructed the distance matrices for store to store, store to warehouse and warehouse to store. If new stores are added then the distance matrix algorithm will be triggered. By using this information and constraints we are predicting the journey plan to the salesman.

Product order recommendation

R, SQL, Tableau

Product order recommendation is used to predict the order recommendation of each product(SKU) to the outlets based on the Historical Transactions, Store Check, Promotions, Core-SKU , Missed Orders and demographic details of outlet. It is assumed that the data captured will define the purchasing patterns of a store. Demographic and transactions data will be used to identify similar outlets, understand the purchasing patterns of these stores and predict recommended order for the further visits to store.

Document Layout Analysis

Opencv,Python

In this project, I did a "proof of concept" on dividing the newspaper into images and text regions. I used the concept of connected components for the extraction of components from the news paper. Here I used a custom distance metric to calculate the distance between adjacent components present in the news paper.

Extracting tabular data from PDF documents

Python

In this project, I did a "proof of concept" for identifying the tables in the PDF's and extracting the information from the tables. I used k-means algorithm for extracting the column data and arranged the data according to the row, based on the word boundary information.

Major Projects

Credit scoring model using German dataset

KNIME,R,Python

Credit scoring model is a process used by credit providers to check the worthiness of the customer and decide whether to provide credit to customers. Credit scoring is a challenge due to increase in credit industry where credit providers face problems to handle the large amount of credit data. To deal with data, several data mining techniques can be used. Since it is a classification problem to predict the customer as good or bad in order to provide credit, data mining classification methods are applied. Some of the classification methods are decision trees, support vector machines, multilayer perceptron, logistic regression etc. The aim of this project is to apply different classification techniques on training set of german credit dataset and come up with different models which give the best predictive results on the test instances.

Association analysis on Groceries dataset

KNIME,R,Python

The goal of this project is to perform Market Basket analysis on Groceries dataset by finding association rules on purchased items. This information will enable the retailers to understand the buyer's needs and rewrite the stores layout accordingly, to develop cross promotion program, or even to capture new buyers. Apriori and Fp Growth algorithms were used to find the association between items. The evaluation metric used to find the best association rules are Support, Confidence, Conviction, Lift, and Leverage.

Implementation of Exact sentence breaking algorithm using Stanford Core NLP jar file

Java,stanford NLP Jar Files — The goal of this algorithm is to implement an exact sentence breaking on a text file. We downloaded the jar file and then we used it to form our task. This algorithm is used to break any text file into exact sentences considering all rules and constraints to form a sentence.

Recommender system to recommend movies

Matlab, Matrix Factorization — The goal of this project is to build a recommendation engine so that we can recommend new movies to the users based on their tastes and preferences. In this project I used Matrix Factorization technique for predicting the ratings of new movies or unwatched movies for the user's based on their past seen movies list. I used precision and recall methods for evaluating the recommender engine.

Probabilistic Matrix Factorization in Cross Domain

Matlab, Kmeans, MF — Traditional Recommender Systems suggest items belonging to single domain. Nowadays users provide feedback for items of different types and express their opinions on different social media, wish to cross-sell products and services and want to provide recommendations to new users. By using cross domain recommender systems we can transfer knowledge from source domain to target domain. Cross domain recommendation addresses the cold start problem, improves accuracy and adds value to the recommendations. In my project I used probabilistic Matrix Factorization in cross domain using k-means clustering and achieved good precision and recall compared to single domain recommender systems.

Technical Skills

Languages	C, R, Python
IDE/DB	Ipython Notebook,RStudio,PyCharm
Machine Learning/Deep Learning Tools	KNIME,Tableau,Keras,NLTK,Scikit Learn

Relevant Coursework in IIITH

Statistical Methods in AI
Natural Language Applications
Computational Linguistics
Database Systems
Distributed Systems

Achievements

Funding	**Funding for the NIPS-2020 conference**.
Subreviewer	**Subreviewer for the EACL-2021 conference**.
Subreviewer	**Subreviewer for the EMNLP-2020 conference**.
Subreviewer	**Subreviewer for the ICON-2019, 2020 conferences**.
Publication	**Research paper in NIPS, ECML, PACLIC and KDD and NIPS Conferences.**.
Certificate	Introduction to Data Science in Python Course in Coursera.
Gate	Secured 81.64% in GATE-2014 Examination in CSE.