YASH SHREESH DUBEY

Arlington, TX, USA • +1(682)407-6794 • yashd995@gmail.com
linkedin.com/in/yash-dubeycs/ • github.com/yashdubey95

Education

Master of Science in Computer Science [GPA: 3.5/4.0] *Aug 2018 - May 2020(Expected)*
The University of Texas at Arlington, Arlington, TX

- Specializing in ***Intelligent Systems (Artificial Intelligence) and Databases***
- Relevant Courses: Artificial Intelligence I, Computer Vision, Neural Networks, Data Mining, Big Data Analytics, Data Analysis and Modeling Techniques, Design and Analysis and Algorithms, Software Testing, Information Security I.

Bachelor of Engineering in Computer Science and Engineering *May 2013 - May 2017*
Dr. Babasaheb Ambedkar College of Engineering and Research, Nagpur, IN

- Course Highlight: Artificial Intelligence, Data Structures, Compiler Design, Computer Architecture, Object-Oriented Programming, DBMS, Software Engineering and Project Management.

Skills

Programming Languages: Proficient in Python, SQL, C/C++; Familiar with R, Java, HTML/CSS.
Data Science Libraries and Frameworks: Pandas, NumPy, Scikit-learn, Matplotlib, TensorFlow, Pytorch, NLTK, ggplot2, mlr, Apache Hadoop, Apache Spark (PySpark).
Machine Learning: Classification, Regression, Clustering, Natural Language Processing, Deep Learning.
Soft Skills: Teamwork, Detail Oriented, Adaptability, Analytical Thinking, Communication, Active Learning
Hard Skills: Data Analysis, Data Visualization, Shell Scripting, Probability & Statistics, Data Structures & Algorithms

Academic Projects

Self-Driving Car *May 2019 – Jul 2019*

- Built a ***Q-learning model*** over a ***Deep Neural Network*** architecture to implement a Modelled Version of a Self- Driving Car that can successfully navigate itself in an environment.
- Concepts implemented to achieve this were ***Reinforcement Learning, Experience Replay, Action Selection Policies***.

Chat Bot *Feb 2019 – Jun 2019*

- Designed and Trained a ***Deep NLP model*** on a ***Seq2Seq Architecture*** to create a chatbot using the ***TensorFlow RNN libraries***.
- The dataset used to train the chatbot is the ***Cornell Movie Corpus Dataset***, which contains nearly ***220,000 lines of conversation*** between characters from over ***600 movies***, this helps in creating a chatbot that can have ***general conversations***.

Predicting NYT's Pick from a pool of Comments *Mar 2019 – May 2019*

- Implemented a ***Logistic Regression model*** to predict whether a comment posted on an article in the New York Times will be picked up the Editor as the best comment i.e. ***NYT's Pick***.
- The dataset had over ***2 million rows and 47 columns***, used data pre-processing techniques such as tokenization, stemming, ***bag-of-words technique*** to eliminate stop words, punctuations, NER, and further vectorizing them using ***TF-IDF*** scores to achieve an ***accuracy of 85%***. The same was achieved using ***PySpark (pyspark.ml library)***.

Movie Recommender System *Oct 2018 – Jan 2019*

- Implemented a ***Content-Based Recommender System*** that computes a pair-wise similarity for all the movies taken from the ***IMDb's Top 250 Movies*** data set based on the lead actors and actresses, the director, the plot of the movie and the genres to which the movie belongs to.
- It then recommends the movies based on the ***Cosine Similarity*** scores taken from the ***Count Vectorizer***.

Topic Detection by Clustering Keyword *Dec 2016 – May 2017*

- Designed a web application that takes an article as an input, using concepts like ***Natural Language Processing*** (NLP) and ***K-Means Clustering Algorithm*** to provide a ***Document Summarization*** in the form of a title, brief explanation on the topic and summarized bullet points as result to better understand the article.

Research Paper Published

Topic Detection by Clustering and Text Mining: International Research Journal of Engineering and Technology, Volume 4, Issue 3, March 2017.

Additional Section

- Completed Certification Courses on: "***Neural Networks and Deep Learning by Andrew Ng***"(Coursera), "***Artificial Intelligence A-Z: Learn how to build an AI***"(Udemy), "***Machine Learning A-Z: Hands-on Python & R in Data Science***" (Udemy), "***Deep Learning A-Z: Hands-on Artificial Neural Networks***"(Udemy).
- Secured **1st position** in "***Technobuzz***" (A national level coding competition).
- Awarded Certificate of Appreciation for Participating in Inter-College and National Level Conventions.
- Team head and Part of Organizing Committee for various Technical and Non-technical events conducted in College.