YASH SHREESH DUBEY

yashdubey95@outlook.com | +1(682)407-6794 | linkedin.com/in/yash-dubeycs/ | github.com/yashdubey95 | yashdubey95.github.io

Education

Master of Science, Computer Science   *Aug 2018 - May 2020*
The University of Texas at Arlington   **GPA: 3.42**
- Specializing in **Intelligent Systems (Artificial Intelligence) and Database**.
- Coursework: Artificial Intelligence I, Machine Learning, Computer Vision, Neural Networks, Data Mining, Big Data Analytics, Data Analysis & Modeling Techniques, Design & Analysis of Algorithms, Software Testing, Distributed Systems.

Bachelor of Engineering, Computer Science and Engineering   *Jun 2013 - May 2017*
RTM Nagpur University
- Coursework: Artificial Intelligence, Data Structures, Compiler Design, Computer Architecture, Object-Oriented Programming, Database Management System, Software Engineering and Project Management.

Work Experience

Research Assistant, The University of Texas at Arlington   *Jul 2020 - Present*
- Working for The Vision-Learning-Mining (VLM) Lab under the guidance of Professor Vassilis Athitsos.
- Currently pursuing research to solve complex computer vision problems like "3D Hand Pose Estimation" and "Assessing Cognitive skills in Children through Performance in Physical and Computer-based Tasks".
- Developing an algorithm for the depth estimation task used to create a 3D orientation of the given input hand shape using PyTorch.
- Analyzed large datasets and worked with raw sensor data by performing data collection, data annotation and data preprocessing tasks.

Project Experience

Monocular Depth Estimation *[Python, Tensorflow]*   *Oct 2019 – Jan 2020*
- Designed and Implemented an **Encoder-Decoder style CNN architecture** to create a supervised-learning model; that took as an input a Single RGB image taken from **a Monocular Source** and created a **Representation of its Spatial Structure** to give a **Depth Map** as an output.

Self-Driving Car *[Python, PyTorch, Kivy]*   *May 2019 – July 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 *[Python, Tensorflow]*   *Feb 2019 – Jun 2019*
- Designed and Trained a **Deep NLP model** on a **Seq2Seq Architecture** to create a chatbot using the **RNN (LSTM) model**.
- 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 *[Python, Scikit-learn, PySpark]*   *Mar 2019 – May 2019*
- Implemented a **Logistic Regression model** to predict whether the Editor will pick a comment posted on an article in the New York Times as the best comment i.e. **NYTs 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 *[Python, NLTK, Scikit-learn]*   *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**.

Skills

Programming Languages: Python, JavaScript, Java, HTML/CSS, SQL, R, C/C++
Data Science Libraries and Frameworks: Pandas, NumPy, Scikit-learn, Matplotlib, OpenCV, TensorFlow, PyTorch, NLTK, ggplot2, mlr, Apache Hadoop, Apache Spark (PySpark), Flask, SciPy, Git
Cloud Platforms: IBM Cloud, Microsoft Azure, Amazon AWS, Google Cloud Platform (GCP)
Machine Learning & Deep Learning Algorithms: Logistic Regression, Linear Regression, Naïve-Bayes, Decision Tree, Random Forest, K-NN, K-Means, SVM, PCA, CNN, RNN
Hard Skills: Data Analysis & Visualization, Shell Scripting, Probability & Statistics, Data Structures & Algorithms.
Soft Skills: Teamwork, Detail Oriented, Adaptability, Analytical Thinking, Communication, Active Learning.

Research Paper Published

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

Certifications & Awards

- Completed a 5 Course **Specialization in Deep Learning** taught by **Prof. Andrew Ng** offered by **deeplearning.ai**.
- Completed certification courses on: "**Artificial Intelligence A-Z: Learn how to build an AI**"(Udemy), "**Machine Learning A-Z: Hands-on Python & R in Data Science**" (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.