JITESH PABLA

Tempe, AZ • 480-930-2008 • jpabla1@asu.edu • jiteshpabla.github.io • linkedin.com/in/jiteshpabla

EDUCATION

Master of Science - Computer Science   **Aug 2019 – May 2021**
Arizona State University, Tempe, AZ   GPA: 3.91/4.0
Courses: Data Mining, Natural Language Processing, Statistical Machine Learning, Data Visualization, Distributed Database Systems

Bachelor of Technology (with honors) - Computer Science and Engineering   **Jul 2015 – May 2019**
Jaypee Institute of Information Technology, Noida, India   GPA: 3.6/4.0
Courses: Deep Learning, Artificial Intelligence, Data structures, Algorithms, Probability, Linear Algebra, Statistics

TECHNICAL SKILLS

Languages and Frameworks: Python (NumPy, Pandas, Scikit-learn, Matplotlib, Keras, PyTorch, TensorFlow, NLTK, OpenCV, Flask, PySpark), JavaScript (JQuery, Vue.js, D3.js), R (tidyverse, ggplot2), SQL/PostgreSQL, Lua, Java, PHP, C++, C
Tools and Miscellaneous Technologies: Git, Google Colab, Jupyter Notebook, Agile, AWS (EC2, S3), Tableau, Hadoop, Spark, MongoDB, MS Excel, Docker, Airflow, A/B Testing, HTML, CSS, Adobe XD, Alteryx; Linux, Windows, macOS;
Certifications: Deep learning specialization - deeplearning.ai (Link)

WORK EXPERIENCE

Data Engineer and Web Developer, Arizona State University, USA   **Nov 2020 – May 2021**
- Migrated the data of over 28 websites from Drupal 7 to Drupal 9 with migration tools by creating ETL pipelines and utilizing SQL to understand and manipulate the large database.
- Designed and built a website - crimeandjusticenews.asu.edu by applying the latest ASU web standards and front-end design.
- Managed the team's kanban board to deliver results on time and increase work efficiency by as much as 10%.

Graduate Research Assistant - Machine Learning, Arizona State University and Mayo Clinic, USA   **Jan 2020 – May 2020**
- Classified 50k COVID-19 articles related to vaccines and therapeutics by scraping Google search results to obtain noisy data and training a scientific-text-based Bidirectional Encoder Representations from Transformers (SciBERT) model.
- Ranked COVID-19 articles for queries relevant to vaccines and therapeutics by utilizing BERT as an embedding generator and finding each article's Cosine similarity with keywords related to vaccines and therapeutics.
- Identified Randomized Controlled Trials from over 50k highly imbalanced PubMed articles by modifying the BERT architecture and manipulating its inputs along with various NLP techniques using PyTorch and transformers.

Software Engineer, Google Summer of Code 2018 Participant with LuaRocks, Remote   **Jun 2018 – Aug 2018**
- Refactored the core functionalities of the LuaRocks commands for - listing, uninstalling, showing details of packages, searching and installing rocks from the web, opening documentation, etc., to modularize them.
- Programmed a complete API to provide access to the LuaRocks functionality using Object-Oriented design patterns and used Git extensively for contributing to the main code-base.
- Designed a responsive and interactive web-based GUI using HTML, CSS, Bulma, and Vue.js to access the LuaRocks functionality.

Data Scientist Intern, Team Computers Pvt. Ltd., India   **Jun 2017 – Jul 2017**
- Applied data preprocessing, machine learning, and statistical methods such as moving averages, linear regression, spectral clustering, etc., on dummy datasets using Alteryx.
- Predicted prospective car customers using car sales and inquiry data (with millions of data points spanning across one year) using time series analysis to potentially boost sales for multiple car dealerships by up to 19%.

PROJECTS

Data Driven Disaster Response (code)   **Aug 2020 – Dec 2020**
- Led a team of six people by organizing meetings, delegating work, and tracking tasks via a kanban board to design an interactive D3.js based dashboard for visualizing a city's social media data to aid the disaster response during a natural disaster.
- Cleaned and Categorized the social media messages into resource categories using statistical metrics and Latent Dirichlet Allocation (LDA) and applied rule-based sentiment analysis using NLTK.
- Developed a set of interconnected visualizations, including - line charts, pie charts, heat maps, etc., to view the frequency of a resource need or a particular emotion in any part of the city during any time.

Clinical Semantic Textual Similarity (code)   **Aug 2019 – Dec 2019**
- Preprocessed the clinical text to remove stop words, punctuation, etc., and utilized various word2vec pre-trained models to extract token embeddings to create a single vector representation for each sentence.
- Fine-tuned multiple BERT models on the given STS dataset and extracted vector representation for each sentence.
- Engineered several similarity features based on the extracted sentence vectors, applied gradient boosting regression, and grid search to achieve a Pearson correlation greater than 0.84 between the ground truth and the model's predictions.