Shruti Padole

812-778-6792 | spadole@iu.edu | LinkedIn

EDUCATION

Indiana University, Bloomington   Bloomington, IN
Master of Science in Computer Science (GPA 3.5)   *Aug. 2021 – May 2023*

CDAC-ACTS   Pune, MH, India
Post Graduate Diploma in Big Data Analytics (GPA 3.8)   *Aug. 2017 – Feb. 2018*

Sinhgad College of Engineering   Pune, MH, India
Bachelor of Engineering in Information Technology (GPA 3.6)   *Jul. 2013 – May 2017*

EXPERIENCE

Social Science Consulting LLC   Bloomington, IN
Data Scientist Intern   *Aug. 2023 – Present*
- Engineered an interactive online resource library for race and religion research, seamlessly integrating **Node.js** and **Python**. Executed **Excel-to-MySQL** migration, driving a 25% surge in user engagement through an **intuitive UI**.

ZS Associates   Pune, MH, India
Data Engineer   *Apr. 2020 – Jul. 2021*
- Designed a pharmaceutical app, integrating **SQL, Spark,** and **AWS**. Streamlined big data processing from Salesforce and Veeva Systems, utilizing **Parquet tables** for quick queries. Leveraged **SparkSQL**, **AWS EC2** deployment, and **S3** for backup and storage.
- Utilized **Python** and **Bash** scripting, and **API** integration to automate weekly backups, data storage in timestamped **AWS-S3** buckets, weekend EC2 instance shutdowns, resulting in elimination of manual labor.
- Utilized **Azkaban Scheduler** adeptly to execute extensive Salesforce file ingestion and data manipulation via **Hadoop** jobs, effectively eradicating manual labor and realizing a remarkable 50% reduction in task completion time.

E42 Light Information Systems   Pune, MH, India
NLP Engineer   *Feb. 2018 – Feb. 2020*
- Spearheaded a team of 8 to develop a robust, scalable, and platform-independent chat bot by leveraging **Python, MySQL, Python-Django, REST APIs**, and training it on millions of WhatsApp threads.
- Enhanced the chat-bot's accuracy by 10% by incorporating a pre-trained **BERT** model alongside a rule-based approach.
- Innovated a fault tolerant chat-bot based on Finite State Machines, resulting in a significant 40% increase in customer engagement.

PROJECTS

Trend Analysis and Hate Speech Detection in Tweets   Aug, 2021 – May 2022
GCP, Python, NLTK, SpaCy, NumPy, Pandas, Tweepy, Git, matplotlib
- Extracted over 60,000 tweets from top US news channels in the last two years and used **LDA and clustering** to map top three trends for every quarter. Used **Tweepy API** to pull tweets and **matplotlib** to display the trending news.
- Implemented hate speech detection in 2 years' worth of tweets collected using Tweepy. Employed **LSTM, Bi-LSTM, and K-means clustering** to classify tweets as containing hate speech or being neutral.
- Leveraged **GCP's Compute Engine** to harness high computational power, while enhancing hate speech detection through sentiment analysis and pre-trained NLP models like **BERT**, **DistilBERT**, etc.

Python Date Extractor   Aug. 2019 – Dec. 2019
Python, Git, PyPi
- Developed a **Python module** following **PyPi standards** to seamlessly **extract dates** from diverse text inputs, such as "the day after next Friday," "last year," and "a week ago." Employed Natural Language Processing techniques to effectively identify temporal patterns, enhancing date recognition accuracy by over 85%.

TECHNICAL SKILLS

Languages: Python, PySpark, Java, Scala, C++, R, Shell Scripting, HTML, CSS, PHP, JavaScript, ReactJS, Linux, Vim
Databases: MySQL, Hive QL, Spark SQL, MariaDb, PostgreSQL, MongoDb, ERDPlus, Redis
Dev Tools: Git Version Control, Jira, PyPi, Android Studio, Jupyter, Eclipse, AWS (S3, EC2), Google Cloud, Tableau, Power BI
Libraries: NumPy, Pandas, Scikit-learn, SciPy, NLTK, TensorFlow, Pytorch, Matplotlib, BeautifulSoup, Regex, Pdb