JITESH PABLA

Las Vegas, NV • +1-480-930-2008 • jiteshpabla@gmail.com • linkedin.com/in/jiteshpabla • jiteshpabla.github.io

EDUCATION

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

Bachelor of Technology (with honors) - Computer Science and Engineering   **Jul 2015 – May 2019**
Jaypee Institute of Information Technology, Noida, India   CGPA: 8.1/10.0
Courses: Data structures, Algorithms, Probability, Linear Algebra, Statistics, Deep Learning, Artificial Intelligence

TECHNICAL SKILLS

Languages and Frameworks: Python (PySpark, NumPy, Pandas, Scikit-learn, Matplotlib, Keras, PyTorch), Scala (Spark), SQL/PostgreSQL/MySQL, JavaScript (JQuery, Vue.js, D3.js), Lua, Java, C++, C
Tools and Miscellaneous Technologies: Databricks, Git, GitHub, Azure DevOps, Hive, Docker, Jupyter Notebook, JIRA, MS Excel, PowerBI, SSIS Package, Teradata, Alteryx, Bash, Linux
Concepts: Agile development, Data Quality, Data Analysis, ETL (Extraction, Transformation & Loading)
Certifications: Azure Fundamentals (Link), Deep learning specialization - deeplearning.ai (Link)

WORK EXPERIENCE

Data Engineer, MGM Resorts International, USA   **Aug 2021 – Present**
- Transformed legacy SQL processes from Teradata and SSIS packages into Apache Spark, enhancing the Customer Data Platform (CDP) to support over 100 million unique customers.
- Planned and integrated new attributes for customer data into CDP from various MGM systems by coordinating with stakeholders and translating requirements into epics and user stories, improving data completeness.
- Improved the CDP codebase by developing new features and resolving bugs in multiple Spark processes and provided comprehensive documentation, enhancing team knowledge and productivity.
- Pioneered an automated data cleanup and vacuuming policy across multiple Azure storage types (DBFS, ADLS, Azure Blob Storage), utilizing Python scripting to reduce MoM storage costs by 40%.
- Improved data quality by creating a PowerBI dashboard to monitor changes and abnormalities in high-impact business data attributes, resulting in preemptive data issue resolutions by up to two days.

Web Developer and Data Engineer, 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 YAML ETL pipelines and utilizing SQL to understand and manipulate the large database.
- Designed and implemented the ASU's crimeandjusticenews.asu.edu website using up-to-date ASU web standards and front-end design.
- Increased team productivity by 10% through efficient Kanban board management.

Machine Learning Research Assistant, 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.
- Boosted prospective car sales by up to 19% by predicting prospective car customers using car sales and inquiry data with millions of data points spanning across one year using time series analysis.