**PRAVALLIKA AVULA**

(+1)5715741074 | [avulapravallika2023@gmail.com](mailto:avulapravallika2023@gmail.com) | [linkedin.com/Pravallika](https://www.linkedin.com/in/pravallika-a-04bb171b7/)

Experienced Data Professional with a master’s degree in data Analytics Engineering and 2 years of experience in machine learning, Data analysis, engineering, visualization, and designing ETL pipelines. Skilled in SQL, R, Python, and AWS for extracting insights and transforming business needs into technical solutions, focusing on data-driven strategies to meet corporate objectives.

**EDUCATION**

**George Mason University | GPA:3.88/4.00 Aug 2022 – May 2024** Master of Science in Data Analytics Engineering Fairfax, VA **Chaitanya Bharathi Institute of Technology | GPA:3.26/4.00 Jul 2016 - Sept 2020** Bachelor of Engineering in Electronics and Communication Hyderabad, India

**TECHNICAL SKILLS**

**Programming:** C, Python, Machine Learning, Scala, Redis, SQL, PL/SQL, RStudio, XML  
**Web Development:** WordPress, HTML5, CSS  
**Databases:** Oracle, MySQL (Relational Databases), MongoDB (NoSQL Database)  
**Libraries:** Tidyverse, dplyr, NumPy, Pandas, SciPy, MatPlot, PyTorch, Seaborn, ggplot, TensorFlow, NLTK, Scikit-learn

**Tools:** Docker, MS Office, AWS, Hadoop, Apache Spark, Visual Studio, Jupiter Notebook, Google Collab, Tableau, PowerBI, Git, Excel, MS Azure DataBricks, Anaconda, SAS

**WORK EXPERIENCE**

**George Mason University Jan 2024-Present**

**Graduate teaching assistant Fairfax, VA**

* Assisting in two **R-based** subjects: “Health Data Challenge” and “Health Information Technology Fund”.
* Providing instructional support, graded assignments, and conducted student consultations.
* Facilitating discussions and assisted students in understanding course material and assignments.

**George Mason University Aug 2023-Dec 2023 Graduate Research Assistant Fairfax, VA**

* Analyzed ADHD and Autism Reddit posts using **Google** **Collab** for **Python**-**based** **topic** **modeling** and created visualizations.
* Conducted topic modeling and sentiment analysis, achieving an 85% accuracy rate using **OpenAI’s ChatGPT API** with **GPT-4** model, to compare shared topics between both communities and analyze frequency and sentiment for each theme.
* Extracted 20 topics from r/ADHD and 18 from r/Autism, each characterized by 10 keywords.
* Identified 10 key themes in ADHD discussions and 8 in Autism discussions.
* Found prominent discussion on diagnosis and treatment, medication, dosage, and side effects, and socio-community aspects.

**Larsen and Toubro Infotech Jan 2021 – Jun 2022**

**Data Analyst Pune, India**

* Collaborated on cleaning, organizing, and analyzing client datasets with advanced SQL in **MySQL** and **MS Excel**, improving data accuracy by 20%.
* Developed **AWS Glue ETL** pipelines, reducing processing time by 30%. Utilized **Hadoop, Apache Spark, AWS** products like **RDS, Redshift, S3, Glue, and EC2** for efficient data management.
* Implemented **Tableau** and **PowerBI** visualizations for data-driven strategies increasing customer engagement.
* Conducted database query tuning, enhancing system efficiency by 40% using **MySQL Workbench, Hadoop, Redis.**
* Integrated machine learning algorithms, reducing client operational costs by 25%.

**Electronics Corporation of India Limited Jun 2018-Jul 2018**

**Intern Hyderabad, India**

* Designed and implemented an innovative **IoT**-based Alarm System for smart homes using **Arduino**, detecting fire and water levels, resulting in a 30% reduction in security system costs.
* Demonstrated system functionality virtually with **Verilog** and **MATLAB**, achieving a 95% accuracy rate in simulated emergency scenarios.

**PROJECTS**

**Generating Synthetic flight tracks using Generative adversarial networks [Python, Tableau, AWS] Jan 2024-Present**

* Demonstrating GANs' application with Python and AWS to generate >1M synthetic flight tracks by training a Neural network on historic flight tracks in collaboration with the Center for Air Transportation Systems Research (CATSR).
* Revolutionizing airspace design by addressing limitations of historic and simulated flight tracks with more accurate, up-to-date data.

**Wine Quality-Applied Statistics and Visualizations for Analytics [Excel, SQL] Sept 2022-Dec 2022**

* Implemented Regression models in RStudio with a 92.7% accuracy for wine quality prediction and preferences.
* Collaborated with Python libraries like **Matplotlib**, **Seaborn**, and **NumPy** for effective data visualization and preprocessing, resulting in a 25% reduction in processing time.

**Covid-19 Update’s website [AWS, Redshift, Glue, S3, Tableau, WordPress] Mar 2021-Apr 2021**

* Developed COVID-19 dashboard using **WordPress** with interactive forms, optimizing event planning through data synthesis, including individual and family experiences.
* Used **AWS** for seamless storage and processing, achieving a 30% improvement in data efficiency. Used **Tableau** and **Quicksight**, achieving a 20% efficiency boost in regional and demographic trend analysis.