

DEEPIKA ALAGIRISWAMY PANNEERSELVAM

Chicago, IL | deepikapanneerselvam6@gmail.com | https://linkedin.com/in/deepika_panneerselvam/ | <https://github.com/DeepikaPanneer> | +1 (630) 873 0087

EDUCATION

Illinois Institute of Technology, Chicago, IL

August 2023 - May 2025

Master of Data Science, GPA: 3.90

Related Coursework: Data Preparation and Analysis, Big Data Technologies, Machine Learning, Regression, Advance Database Organization, Monte Carlo Methods in Finance, Statistical Learning.

Coimbatore Institute of Technology, Coimbatore, India

August 2017 - May 2021

Bachelor of Technology, Information Technology, GPA: 3.8

Related Coursework: Data Structures, Database Management System, Object Oriented Programming, C Programming, Cloud Computing, Operating Systems, Data Warehousing and Data Mining.

EXPERIENCE

Urban Institute - Washington DC, USA

Data Engineering Intern

June 2024 – November 2024

- Upgraded outdated RStudio, R, and EMR in a 7-8-year-old system by automating the version update process through API calls, seamlessly integrating updates into the existing system.
- Designed and implemented an AWS Step Function setup to efficiently orchestrate workflows, including EMR cluster creation and EC2 instance provisioning, enhancing process automation, scalability, and reliability.
- Created a CI/CD setup to streamline system updates and deployments using AWS Serverless Application Model (SAM) and GitHub Actions, enabling automated and efficient development.

Dun and Bradstreet - Hyderabad, India

Big Data Engineer

August 2021 - July 2023

- Enhanced 80% of datasets by integrating key features, significantly improving infrastructure performance.
- Managed to deploy 3 new data sources into the production environment using AWS, PySpark and Snowflake.
- Created scalable, robust databases that efficiently manage ETL operations using Spark and SQL integration.
- Fixed various high-priority bugs, which reduced the number of failures in production by 30%.

Data Engineer Intern

February 2021 – August 2021

- Documented workflow of various modules and addressed high-priority tasks while deploying to higher environments.
- Monitored critical production jobs, promptly identifying and swiftly resolving any encountered bugs or failures to ensure seamless operation and minimal downtime.

SKILLS

Programming Languages and Libraries:

Python, PySpark, Pandas, NumPy, R, C.

Big Data Technologies and Databases:

Spark, Hadoop, Hive, Snowflake, MySQL, SQL, MongoDB.

Cloud Services:

AWS - EMR, S3, Lambda, SAM, Redshift, Glue, Kinesis, DynamoDB.

Automation and Version Control:

Airflow, Jenkins, Git, GitHub, GitHub Actions, Jira

Data Analysis and Visualization:

MS Excel, PowerBI.

CERTIFICATIONS

AWS Certified Data Engineer– Associate

Amazon Web Services (AWS) – August 2024

PROJECTS

Multilabel Predictions on Academic Articles

April 2024

- Designed and implemented a predictive model for multilabel classification of academic articles, streamlining content categorization across five distinct labels and enhancing retrieval system efficiency.
- Utilized advanced Machine Learning and Natural Language Processing (NLP) techniques, leveraging Python, Pandas, NumPy, and BERT to achieve accurate and scalable classification.

Predictions for the 2023 ICC Cricket World Cup

November 2023

- Developed machine learning models to predict match outcomes in the 2023 ICC Cricket World Cup, leveraging historical data from 10 teams. Conducted in-depth analysis of team performance metrics and contextual factors, and visualized insights using advanced data visualization techniques.
- Utilized technologies such as R for data analysis, Random Forest for predictive modeling, and ggplot2 and Tidyverse for data visualization and manipulation.