# ARTHKUMAR PATEL

**2646 Beacon Hill CT Apt 208, Auburn Hills, Michigan - 48326**

**(949) 527-8145 , arth2766@gmail.com**

Portfolio | LinkedIn | Github

## EDUCATION

**San Diego State University** — San Diego, California
*Master of Science, Computer Science - GPA: 3.83/4.0* — *January 2021 - December 2022*

**LDRP - Institute of Technology & Research** — Gandhinagar, Gujarat, INDIA
*Bachelor of Engineering, Computer Engineering - GPA: 8.58/10.0* — *August 2016 - June 2020*

## SKILLS

- **Programming Languages -** Python, C++, JavaScript, Shell Scripting (Linux & UNIX)
- **ML & Deep Learning Frameworks -** Tensorflow, Scikit-learn, PyTorch, NLP
- **Data Analysis & Visualization -** Numpy, Pandas, Tableau, Microsoft PowerBI, Jupyter Notebook
- **Web Frameworks & Databases –** HTML5, CSS3, React.JS, REST APIs, SQL Server, MySQL, NoSQL – MongoDB, Hive
- **Other –** Github, AWS EMR, Docker, Apache Spark, BitBucket, PySpark, Statistics

## WORK EXPERIENCE

**General Motors** | Python, SQL, Hive, Oracle Database, Power BI, Hadoop, Azure — Warren, Michigan
***IT Data Analyst*** — *January 2023 - Present*

- As a part of Vehicle Applications Integration team, support ad-hoc data requests by performing complex database queries across data sources such as Hive, Oracle SQL Developer and Microsoft SQL Server to extract data needed for solving production tickets.
- Also, works on Data Cleaning, Data Migration and reporting dashboards from MSBI Server and create Power BI Dashboards with interactive visualization reports by modeling data. Also, contributes in creating Python scripts for automating Power BI visuals and provides datasets for Alpha/Beta testing by identifying data patterns across datasets of the vehicle applications.

**RipeMetrics** | Python, SQL, Apache Spark, ML, Airflow, MongoDB, AWS S3, Sagemaker — San Diego, California
***Data Science Intern*** — *January 2022 - May 2022*

- Performed Data Preprocessing & Deep Exploratory Analysis using Python, Pandas, Numpy and created effective and interactive data visualization charts and dashboards in Tableau. Structured data queries & data mining through MongoDB Database.
- Designed Recommendation and Predictive ML Models using Scikit-learn, PyTorch and Sagemaker. Built data pipelines and batch jobs using Python & Airflow. Collaborated in big data extraction & processing of real-time streaming data using Apache Spark, AWS S3 and EC2.

**SDSU Global Campus** | Python, AWS Athena, Redshift, MySQL, Tableau, Hadoop — San Diego, California
***Data Analysis & Digital Marketing Assistant*** — *July 2021 - January 2022*

- Performed Data analysis using Python libraries, Database Monitoring & Management using MySQL Relational Database on the digital marketing campaigns data & visualized it using Tableau dashboards to compare different Key Performance Indicators (KPIs).
- Identified patterns, trends and relationships within data to find business insights. Implemented & supported reporting and analytics infrastructure for customers using AWS services like Athena and Redshift.

**WaytoWeb Pvt. Ltd.** | Jenkins, REST APIs, Git, Linux, Agile, C++, JIRA, Docker, Kubernetes — Ahmedabad, Gujarat, INDIA
***Software Engineer Intern*** — *June 2019 - November 2019*

- Identified and translated customer requirements into detailed architecture and designed system. Developed test scripts for automation and performed unit testing of REST APIs and other services using C++.
- Ensured Unit, Integration tests, Continuous Integration & Deployment (CI/CD) using Jenkins, Git, Linux and Agile Scrum practices for efficiency. Contributed to developing highly scalable end to end application using HTML5, CSS3 and ReactJS Framework.

## PROJECTS

**NBA Games & Players Data Analysis** | Python, Numpy, Pandas, Matplotlib, Data Analysis, IPython, Data Visualization — **Link**

- By using NBA data between 2003-2020, performed analysis and visualized some of the results using pandas, numpy, matplotlib, seaborn, plotly and interactive python functions. Created players and teams performance Comparison models, answered set of questions about players' performance and teams' performance by my analysis. Built Predictive models like Regression and Classification to predict performance.

**Presidential Election Donations Analysis** | Python, Apache Spark, PySpark, AWS EC2, AWS S3, Numpy, Pandas — **Link**

- From the dataset of more than 30 million donations, I analyzed how much money each candidate received in 2020 US Election, also finding mean, median of donations for each candidate, unique contributors, monitoring small contributors and their impact on Presidential Campaigns. Implemented code locally using Spark Shell and on cloud using AWS EC2 which provided faster implementation of Dataframes using SparkSession from pyspark.sql

## CERTIFICATIONS

- Google Data Analytics Specialization – By Google and Coursera
- IBM Data Science Specialization – By IBM and Coursera
- Stanford University ML Certification – By Stanford University and Coursera
- Udemy Tableau Certification – By Udemy