# Sivaprakasam Kaveya
## Data Scientist

San Diego, CA | +1 (619) 877-4937 | kaveyas912@gmail.com

Linkedin | Github

## SUMMARY

- Around 3 years of experience as a data scientist designing, developing, and deploying machine learning models to solve complex business challenges.
- Knowledge of Software Development life Cycle (SDLC), Agile, and Waterfall methodologies, and proficiency in Python, R, SQL, and SAS for data analysis, modeling, and automation.
- Expertise in applying Linear Regression, Logistic Regression, Decision Trees, SVM, Random Forests, Naive Bayes, KNN, K-means, CNN, and Supervised and Unsupervised Learning techniques for data-driven decision-making and predictive analysis.
- Proficiency in NumPy, Pandas, Matplotlib, SciPy, TensorFlow, and Ggplot2 for data manipulation, visualization, and machine learning model development.
- Experienced in AWS, Azure, and cloud platforms for scalable data solutions, with proficiency in Tableau and Power BI for data visualization and reporting.
- Skilled in SQL Server, MySQL, PostgreSQL, and MongoDB for managing and querying structured and unstructured data.
- Familiar with Git, GitHub, and MS Excel for efficient collaboration, versioning, and data analysis.

## SKILLS

**Methodologies:** SDLC, Agile, Waterfall
**Language:** Python, R, SQL, SAS
**IDEs:** Visual Studio Code, PyCharm
**ML Algorithm:** Linear Regression, Logistic Regression, Decision Trees, Supervised Learning, Unsupervised Learning, SVM, Random Forests, Naive Bayes, KNN, K Means, CNN, Classification

**Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Ggplot2
**Cloud Technologies:** AWS, Azure, GCP, DataBricks
**Visualization Tools:** Tableau, Power BI
**Database:** SQL Server, MySQL, PostgreSQL, MongoDB
**Version Control Tools:** Git, GitHub, MS Excel
**Operating System:** Windows, Linux, Mac

## EDUCATION

**Master of Science in Data Science |** University of California, San Diego

**Bachelor of Engineering in Computer Engineering |** University of Mumbai, India

## EXPERIENCE

**Pfizer, USA |** Data Scientist                                                                                   **Jun 2024 - Present**

- Employed Agile, improving workflow efficiency and reducing project delivery time by 20% in dynamic data environments.
- Built and deployed machine learning models, including Random Forests, Naive Bayes, KNN, K Means, and CNN for classification and clustering tasks.
- Leveraged SciPy, Scikit-learn, Seaborn, and TensorFlow, boosting model performance accuracy by 25% and enhancing data visualization quality.
- Accomplished cloud-based data solutions on Azure, reducing data pipeline latency by 40% and scaling machine learning applications efficiently.
- Created impactful data visualizations and reports with Tableau, increasing decision-making speed across teams by 15%.
- Worked with databases like MySQL and PostgreSQL to handle large datasets, improving query efficiency and data integrity.
- Intellectualized GitHub for version control, improving code integrity and reducing versioning issues by 35%, enabling smoother team collaboration.

**Mindtree, India |** Data Scientist                                                                           **Apr 2021 - Aug 2023**

- Applied Waterfall for structured project management, ensuring successful and timely delivery of data science solutions.
- Used Python and SQL to develop and optimize data models, leading to a 30% increase in data processing speed and automation.
- Industrialized Decision Tree models and applied Supervised and Unsupervised Learning techniques, enhancing accuracy by 35% and reducing prediction errors by 25%.
- Implemented K Means clustering and CNNs using TensorFlow, Scikit-learn, SciPy, and Seaborn, improving data segmentation efficiency and enhancing image classification accuracy for actionable insights.
- Managed large-scale data processing on AWS, optimizing data pipelines and significantly reducing processing times.
- Premeditated interactive Power BI dashboards that streamlined reporting processes and reduced manual analysis time by 45%, providing critical insights to stakeholders.
- Improved data retrieval and management in MongoDB, enhancing query response times and overall database performance.
- Conceptualized Git for version control, ensuring high code integrity and streamlined collaboration across projects.

## CERTIFICATIONS
**Diploma in Data Science - (IITM) Indian Institute of Technology, Madras**