

# Sooraj S

Phone : +91 8122152267 | Email: [sooraj12sreekumar@gmail.com](mailto:sooraj12sreekumar@gmail.com) | LinkedIn : [www.linkedin.com/in/sooraj12sreekumar](https://www.linkedin.com/in/sooraj12sreekumar) |  
Github : <https://github.com/Sooraj12Sreekumar>

## Education

---

### VIT Bhopal University

Int M.Tech

Major in Computer Science and Engineering; Minors in Computational and Data Science

Cumulative GPA: 8.35/10

Bhopal, Madhya Pradesh

Expected May 2026

### 12<sup>th</sup> Standard

D.A.V Matric.Hr.Sec.School.

Percentage : 87.3

Chennai, Tamil Nadu

May 2021

### 10<sup>th</sup> Standard

D.A.V Matric.Hr.Sec.School.

Percentage : 82

Chennai, Tamil Nadu

May 2019

## Technical Skills

---

- **Programming Languages:** Python.
- **Frameworks:** Tensorflow.
- **Data Visualization:** Tableau, Power BI.
- **Database Management:** SQL.
- **Spreadsheet Software:** Excel.
- **Advanced Analytics and AI :** Machine Learning and Computer Vision.

## Projects

---

### Resume Screening

June 2024 – Aug 2024

- Developed and evaluated five machine learning models (Random Forest, SVM, Naive Bayes, Logistic Regression, and KNN) for automated resume screening, achieving up to 99.5% accuracy in classifying qualified candidates.
- Implemented text preprocessing techniques, including tokenization, lemmatization, and TF-IDF, to optimize resume data for machine learning algorithms, resulting in improved model accuracy.
- Achieved up to 99.5% accuracy in classifying qualified candidates through the development and evaluation of five machine learning models (Random Forest, SVM, Naive Bayes, Logistic Regression, and KNN) and the implementation of text preprocessing techniques, including tokenization, lemmatization, and TF-IDF.

### Electricity Consumption Forecasting

May 2023 – July 2023

- Applied time series analysis and ARIMA modeling to forecast residential electricity consumption, identifying key seasonal trends and informing potential energy efficiency initiatives.
- Utilized statistical methods, including the Augmented Dickey-Fuller Test (ADF) and rolling mean/standard deviation analysis, to ensure data stationarity, a critical requirement for accurate ARIMA modeling.
- Identified key seasonal trends and informed potential energy efficiency initiatives by applying time series analysis and ARIMA modeling to forecast residential electricity consumption, utilizing statistical methods, including the Augmented Dickey-Fuller Test (ADF) and rolling mean/standard deviation analysis, to ensure data stationarity, a critical requirement for accurate ARIMA modeling.

### Lung Cancer Prediction

Aug 2022 – Oct 2022

- Developed a machine learning model using Python to predict lung cancer with 91.26% accuracy leveraging various algorithms to analyze patient attributes for early disease detection.
- Analyzed patient attributes from a Data World dataset using Pandas, Matplotlib, and Scikit-learn to identify predictive factors for early lung cancer detection.

- Identified predictive factors for early lung cancer detection by analyzing patient attributes from a Data World dataset using Pandas, Matplotlib, and Scikit-learn, leading to the development of a Python-based machine learning model with 91.26% accuracy in predicting lung cancer.

## Experience

---

### Intern at Kumaran Systems-Chennai.

Sept 2024 – Jan 2025

- Successfully applied Optical Character Recognition (OCR) techniques to extract and analyze text from diverse image datasets. Developed scripts and workflows to automate the OCR process, significantly improving data extraction efficiency.
- Played a key role in annotating and preparing large-scale image datasets for machine learning projects. Utilized tools and techniques to ensure accurate labeling of data, facilitating the training of computer vision models.

## Certifications & Training:

---

- Applied Machine Learning in Python- Coursera(University Of Michigan), JAN 23.
- The DataScientist's ToolBox- Coursera(Johns Hopkins University), DEC 22.
- Statistics for Data Science and Business Analysis- Udemy, AUG 23.

## Extracurricular

---

### Data Science Club

Nov 2021- Oct 2022

- Member of the Data Science Club at VIT Bhopal

## Co-Curricular

---

- Five star rank in Hackerrank for SQL.
- Solved more than 100 DSA questions on both Leetcode and GeeksforGeeks focusing on arrays and strings.

## Additional

---

**Languages:** Fluent in, English, Malayalam,Tamil.