

# Gopinath Rathinakali

Gopinathrathinakali.com; +91-95000 39401; gopinath.rathinakali@gmail.com

INTEGRITY

RESPECT

PERFORMANCE EXCELLENCE

ACCOUNTABILITY

**Summary (4+ yr's exp as Data Scientist; 6 yr's exp as Data Analytics & Visualization; 5 yr's exp as Full-Stack; 1 yr exp as testing)**

- With over 16 years of experience in Information Technology, I have gained extensive knowledge in various areas such as predictive modelling, exploratory data analysis (EDA), natural language processing (NLP), deep learning (DL), neural networks, design, installation and implementation of multi-tier Business Intelligence systems, and client-server technology. My experience with Microsoft technology and Data Analytics has allowed me to develop a range of skills.
- As a Data Scientist, I have accumulated **4+ years of hands-on experience in analyzing data, designing, interpreting, visualizing, and deploying machine learning algorithms using Python**. In my work, I have utilized various tools such as Spyder and Jupiter to deploy machine learning models. I have also been able to provide executive leadership teams with insights, analytics, reports, and recommendations, which enable effective strategic planning across all business units, distribution channels, and product lines.
- My experience with supervised and unsupervised learning, as well as text classification, has given me a hands-on understanding of these methods. Additionally, I am efficient in understanding business requirements and transforming them into a data model.

## Experience summary:

- **Infosys India Pvt.Ltd.**, Chennai as **Technology Lead (Jan 2019 to till date)**
- Verizon India Pvt.Ltd., Chennai as Senior Analyst (Mar 2011 to Dec 2018)
- SmarTek21 India Pvt.Ltd., Chennai as Analyst (Sep 2009 to Feb 2011)
- PaddySS India Pvt.Ltd., Chennai as Software Engineer (Dec 2006 to May 2009)

## Technical Expertise

Language	: Python, C#, ASP.Net, VB.Net
Visualization	: Qlik sense, Cognos, tableau, OBIEE, d3js, SSRS
Algorithms	: Regression (Linear, Polynomial, Logistic), Decision tree, K-Nearest Neighbors, Random forest, Gradientboosting, K-Means Clustering
Database	: Oracle, SQL Server, My sql, No sql, Hive
Web	: HTML 5, Bootstrap 4, CSS 3, Javascript, XML, ADO.net
Source	
Control	: Git, VS Team System, VSS, Sub Version
IDE	: Anaconda, Visual studio
Process	: Agile, jira

## Academic Profile

- B.Tech Information Technology – Mohamed Sathak Engineering College Kilakarai (Anna University-75%), TamilNadu.
- Diploma in Computer Technology – (DCT -83.6 %), Mohamed Sathak Polytechnic College Kilakarai, TamilNadu.
- 10th standard – (10th- 63.2 %), Syed Ammal Higher Secondary School, Ramanathapuram, TamilNadu.

## Projects and Clients

### Extreme Network: - Discount % and winning a deal prediction

As a data scientist, I developed a Random Forest Regression model using Python and Incorta to predict the optimized discount for winning a deal or opportunity. This model utilized customer and market data to determine the most effective discount rate that could maximize sales or profitability.

The objective was to leverage the power of machine learning to provide sales professionals with valuable insights, including the probability of converting an opportunity into a win and the optimal discount rate. To accomplish this, the

model was trained using historical data at regular intervals, enabling it to predict the discount rate and win probability for open deals with a high degree of accuracy.

My work on this project involved designing, implementing, and refining the model, as well as analyzing and presenting the results to key stakeholders. By utilizing this model, sales teams can make data-driven decisions and optimize their discounting strategies, leading to increased revenue and profitability.

### **Nissan: - Cheque Image processing**

The objective of this project is to automate the process of collecting information from scanned cheque images using the Python programming language and OCR (Optical Character Recognition) tools such as Tesseract and OpenCV. The program is designed to extract key information such as the bank name, branch, date, amount, and MICR code from scanned cheques.

Once the information is extracted, the program generates an Excel file containing this information and then pushes it to a database. This project is aimed at simplifying the manual process of collecting information from scanned cheques and improving the speed and accuracy of the data collection process.

### **Linked in: - Task hours prediction & Ticket Classification**

I have developed a util program which facilitates the calculation of Service Level Agreement (SLA) for each Jira ticket, predicts the expected hours to complete the task, and classifies the internal and external tickets from Jira. The program further calculates the time taken, IT hours, Non-IT hours, total hours, and displays the difference between the Allowed SLA hours and taken hours. These hours are configured in the spreadsheet "SLA Target" and are calculated based on the "Shift Hours" configured in Excel. Additionally, the program generates the flow of "Ticket Transition Stages" along with the hour taken on each stage.

To identify internal and external tickets and predict the expected hours, I have implemented a Random Forest classifier and Linear Regression algorithm, respectively. The Random Forest classifier has been utilized to accurately classify the internal and external tickets while the Linear Regression algorithm has been used to predict the expected hours.

By implementing this util program, I was able to streamline and automate the SLA calculation process for Jira tickets, leading to increased efficiency and accuracy in task completion.

### **Toshiba & Wesco: - NLP - Ticket classification & Topic Modelling**

have created a powerful tool that offers a Support Ticket Analytics solution. This open-source application is specifically designed for analyzing support tickets and can be hosted on client servers or used within the Infosys network. It is a valuable tool that can be used by both support teams and pre-sales teams to take a deeper dive into ticket analysis.

The Ticket Analytics Dashboard has a range of advanced features that make it an incredibly useful tool for support teams. For example, it can leverage Natural Language Processing (NLP) to analyze frequently appearing terms in tickets, providing valuable insights into the most common issues faced by customers. This makes it easier for support teams to prioritize their efforts and ensure that they are addressing the most critical issues first.

In addition to its NLP capabilities, the Ticket Analytics Dashboard also predicts and assigns ticket categories for newly created tickets. This helps to streamline the support process and ensure that tickets are routed to the right teams quickly and efficiently.

Another significant benefit of the Ticket Analytics Dashboard is its ability to conduct root cause analysis. By analyzing the data collected from support tickets, the application can identify recurring issues and help teams to address them proactively. This can reduce the number of support requests and improve customer satisfaction.

To achieve this, we have used TF-IDF, spacy, and LDA text analysis, which has enabled us to perform topic modelling on the data. This allows support teams to understand the root cause of issues and make data-driven decisions on how to address them.

The Ticket Analytics Dashboard has been designed with integration in mind. It can be seamlessly integrated with ServiceNow or any other ticketing tool, providing a comprehensive solution for support teams. This means that support teams can leverage the application's advanced capabilities while still using the ticketing tool that they are familiar with.

Finally, the Ticket Analytics Dashboard can be used in offline mode, providing added flexibility and convenience. This means that support teams can continue to analyze ticket data even when they are not connected to the internet.

Overall, the Ticket Analytics Dashboard is an advanced tool that can help support teams to improve their efficiency, reduce customer support requests, and enhance customer satisfaction. Its advanced capabilities, including NLP analysis,

topic modelling, root cause analysis, and integration with other ticketing tools, make it an invaluable asset for any support team.

### **SSE: - Predict unit consumption and station classification**

have created a program for SSE's datastore using Python, Pandas, and Scikit-learn libraries to achieve the desired results. The BESP Campaign Datastore implementation was necessary to deliver compiled performance and regulatory reporting of SSE BE Gas and Elect customers and target identification for SMART install. This program has several key high-level use cases, including eligibility funnels for SMART opportunities, renewals dashboards for sales round impact, campaign reporting for performance of rollout, and a DOT governance dashboard for OFGEM/BEIS reporting.

To achieve these goals, we utilized advanced data analysis techniques, including classification and regression algorithms. Specifically, we were able to predict which electric stations consume high volumes of utilization and identify the reasons for the high volume. We collected data on electric stations, trained our model with existing data, and used the classification algorithm to identify areas or electric stations with high consumption.

Our program provides significant benefits for SSE's operations, including the ability to identify SMART opportunities and optimize sales rounds impact. By accurately predicting areas of high utilization, SSE can proactively manage their resources to prevent potential issues and optimize their operations. Additionally, the program provides valuable data for regulatory reporting, helping SSE to comply with OFGEM/BEIS reporting requirements.

### **Shipment Delay date prediction: - Predict expected date of delivery**

have developed a predictive model to determine shipment dates based on historical shipment data details. This program calculates the number of delivery days starting from the source port, transshipment port, and destination port, and then uses regression modelling to estimate the expected delivery date. To achieve this, I implemented various regression models such as RandomForestRegressor, ExtraTreesRegressor, LinearRegression, and DecisionTreeRegressor. After evaluating the performance of each model, we found that the 'RandomForestRegressor' model provided the highest accuracy for predicting shipment dates. By using this program, companies can accurately predict the expected delivery date of shipments, which is essential for maintaining customer satisfaction and optimizing supply chain management

### **Profile screening: - NLP and Talent Classification**

The objective of this project is to automate the process of collecting key information from profiles using two methods. The first method uses the NLP library Spacy and a custom NER (Named Entity Recognition) model to extract information such as name, mobile number, email address, education details, qualifications, experience, and skill sets. The second method utilizes the GPT-3 API to extract similar information.

The program is designed to streamline the manual process of profile analysis and information extraction, improving efficiency and accuracy. By utilizing their expertise in natural language processing and machine learning, the developer of this project demonstrates their ability to develop innovative solutions that automate complex tasks.

### **CompuCom & Citi bank: - Automation**

have created a powerful automation tool that enables organizations to seamlessly set up a fresh instance after acquiring a full-fledged fusion instance. This innovative solution streamlines the process of separating instances and apps, ensuring a smooth transition for acquired companies.

Using Python and Excel configuration, my tool extracts critical configuration data from the existing environment, compares it across instances, and uploads the setup data into the new environment based on the config workbooks. This approach guarantees that the configuration is consistent and accurate, reducing the risk of errors and ensuring a successful setup process.