

Siddharth Srivastava

Experienced data scientist with 5 years of experience and a total of 13 years in IT. Proficient in Python, R, SQL, and Tableau. Skilled in designing and implementing data pipelines, building predictive models, and conducting exploratory data analysis and BI Dashboards.

PROFESSIONAL EXPERIENCE (13 years)

Aquartia Technologies- Data Scientists

Sept 2020- Present

- Define business problems for data analysis.
- Collect, clean, and organize complex datasets.
- Develop statistical models and machine learning algorithms for insights and predictions.
- Communicate findings through visualizations, presentations, and reports.
- Collaborate with cross-functional teams for data-driven decisions.
- Stay updated on industry developments and emerging technologies.
- Transform data science prototypes and design machine learning systems.
- Research and implement ML algorithms, meeting requirements.
- Select datasets and representation methods, run tests, and fine-tune models.
- Extend existing ML libraries and frameworks.
- Analyze large datasets for customer behavior patterns and preferences.
- Develop predictive models for forecasting demand and sales.
- Utilize data mining for optimized marketing campaigns.
- Design and implement data warehouses for large-scale data processing.
- Create automated reports and dashboards for monitoring customer engagement.
- Contribute to refining product development and processes.

Tranzita Systems - Engineering Manager

May 2018- Sept 2020

- Data collection and cleaning by collecting data from various sources, cleaning it, and preparing it for analysis using tools such as SQL and Python.
- Developing and implementing models by developing statistical models and machine learning algorithms to analyze data and make predictions. Implementing models using programming languages such as Python or R.
- Communicating findings by communicating results of data analysis to stakeholders using visualizations, dashboards, and presentations. Providing actionable insights and recommendations to decision-makers.
- Collaborating with cross-functional teams by working with other teams such as product, engineering, and business to understand their needs, and aligning data analysis to achieve common goals.
- Keeping up with new technologies by keeping up with new technologies

Kadubisanhalli
Bangalore-560103

get2sid@gmail.com
Mob: +91-8317066413

TOOLS

Jupyter Notebook

SQL

Microsoft Power BI

Microsoft Power Apps

SKILLS

Python

BI Dashboards

Cloud Computing: Google
Cloud Platform (GCP)

Data Pipelines: Apache
Airflow, Apache Spark,
PySpark

Numpy, Pandas

Statics AI/ML with IOT

Machine Learning: Scikit-learn,
TensorFlow, Keras

Provide concise data reports
and clear data visualizations
for management

Data Visualization: ggplot2,
Plotly, Matplotlib

Natural Language Processing:
NLTK, spaCy

Data Analysis & Modeling:
Python, R, SQL, Tableau, SAS

Mine data from primary and
secondary sources.

Fourier transforms, Wavelet
transform, Kalman filters

DOB:

27-05-1985

INTEREST:

Bike riding, Listening to music,

and trends in data science and machine learning. Understanding new tools, techniques, and methodologies and implementing them in projects where appropriate.

- Ensuring data security and privacy by ensuring the confidentiality, integrity, and availability of data, and adhering to data privacy regulations and guidelines.
- Supervise and lead engineers, plan and develop software projects, and oversee production and quality control.
- Direct and coordinate production, operations, quality assurance, testing, or maintenance in industrial plants.
- Oversee the research and development of new products and procedures.
- Write performance reviews and solve internal issues.
- Discuss and layout project specifications. Make detailed plans to accomplish goals.
- Review, approve, or modify product designs with the management.
- Prepare budgets, bids, and contracts. Negotiate Support, AMC, etc contracts.
- Draft proposals and reports for clients.
- Coordinate activities of the unit with other units or organizations.
- Confer with higher levels of management. Understand user needs and translate them into product requirements.
- Prioritize product requirements based on customer and business impact.
- Own and drive the strategy of products for specific markets and create product roadmaps for those markets.
- Prototyping the product changes and leading the effort of testing the MVP with customers in focused markets.
- The changes I create are incorporated into the product with every release to make the product suitable for global customers.

Aspirants Travel- Sr.SEO Executive

Dec 2015- May 2018

- Building statistical analysis, data visualization, programming languages like Python or R, and machine learning concepts.
- Data collection and analysis by collecting and analyzing data from Google Analytics to understand website traffic, user behavior, and engagement. Utilizing basic analytics techniques to derive insights from the data.
- Tracking and reporting by assisting in setting up tracking mechanisms to capture user behavior on websites and applications. Assisting in creating custom reports and dashboards to monitor performance and inform business decisions.
- Conversion rate optimization: Assisting in identifying key areas of the website or application to improve conversion rates. Implementing changes based on data analysis to drive business growth.
- Business intelligence: Assisting in providing insights to stakeholders about website and application performance, user behavior, and trends. Collaborating with cross-functional teams to develop data-driven solutions for business problems.
- Conduct Keyword research with apt searchers' Intent and competitive analysis. Suggest technical SEO changes regarding website crawl budget, architecture, link building, etc.

- Worked with in-house content writers and freelance writers to create SEO-specific content.
- Research and analyze competitor links and perform link-building outreach via emails.
- Optimize copy and landing pages for organic traffic and growth.
- Execute tests, collect, and analyze data and results, and identify trends and insights in order to achieve maximum ROI in SEO activities.
- Track, report, and analyze website analytics, search console, and ranking reports for different campaigns.
- Work with the development team to ensure SEO best practices are properly implemented on newly developed code.

Career Drudge Technology- SEO Executive

April 2011- Dec 2015

As an SEO Executive I had to handle communicating with clients and creating business strategies and Off Page Optimization: Smart linking strategy, Local & specialized quality business listing searching and submission, Search Engine Submission with creating the business listing.

Lupin Pvt. Ltd (Lucknow) - Marketing Executive

Oct 2009- March 2011

As a marketing executive, I had to visit doctors and retailers to convince them to prescribe and sell company products. Solely helped in scaling sales by 30% in a new territory.

EDUCATIONAL QUALIFICATION

- MCA from Mahatma Gandhi University (MGU) - 2015
- BCA from IMS Engineering College Dasna, Ghaziabad affiliated to Chaudhary Charan Singh University, Meerut - 2009
- 12th from National Institute of Open Schooling.
- 10th from C.B.S.E.

Project Summary

Project: Customer segmentation for a retail chain

Client: Bharatnatyamworld.com

Role: Data Scientist

Technology: Python, R, Scikit-learn, Pandas, NumPy, Matplotlib and Seaborn, SciPy, K-Means Clustering, Decision Trees and Random Forests, Logistic Regression, Tableau.

Project Description: To optimize customer segmentation for BharatanatyamWorld.com, we employ data-driven strategies. By collecting and analyzing customer data, we'll use clustering algorithms to automatically identify distinct segments based on demographics, location, and purchasing patterns. This segmentation enable personalized marketing campaigns and targeted promotions, enhancing customer engagement and satisfaction. Regular adjustments made to adapt to changing behaviors and market trends. A continuous feedback loop and clear reporting validate and enhance the effectiveness of these strategies, fostering business growth for BharatanatyamWorld.com.

Responsibilities:

- Data Collection and Preprocessing: I was responsible for collecting the relevant data related to Bharatanatyamworld.com, and preprocessing the data to make it usable for analysis.
- Data Analysis: I was responsible for analyzing the data to identify patterns and trends. I was using machine learning algorithms and statistical models to identify potential faults or maintenance needs before they occur.
- Model Development: Apply machine learning algorithms, such as clustering models, to segment customers based on predefined criteria.Fine-tune and optimize models for accuracy and performance
- Performance Monitoring: I was responsible for monitoring the performance of the models in real-time, making adjustments as needed to improve accuracy and reduce false positives.
- Reporting: I was responsible for creating reports and dashboards that summarize the health of the model
- Collaboration: Collaborate with cross-functional teams, including marketing, product management, and IT, to ensure the successful integration of segmentation strategies into business operations.

Project Summary

Project: Sentiment analysis for social media data

Client: dhlcargopackersmovers.com

Role: Data Scientist

Technology: Python, R, NLTK,spaCy,Scikit-learn,Pandas,NumPy,TensorFlow,Matplotlib,Seaborn,Tableau,Apache Kafka, Facebook Graph API, Twitter API

Project Description: The project for dhlcargopackersmovers.com focuses on improving online engagement through sentiment analysis on social media platforms. Using advanced NLP and machine learning models, we'll analyze user-generated content to understand sentiments about the brand. The project includes customizing models for industry language, providing real-time analysis, ensuring scalability, and implementing crisis management strategies. Visualizations and reports will summarize sentiment trends, and a continuous feedback loop will refine models based on user feedback. Integration with customer support processes will enable proactive addressing of concerns raised on social media, ultimately enhancing the brand's online reputation and customer satisfaction.

Responsibilities:

- Gather and clean social media data related to DHL Cargo, ensuring accuracy and completeness.
- Conduct EDA to understand data distribution and identify patterns influencing sentiment analysis.
- Select relevant features based on EDA insights.
- Engineer new features if needed to improve sentiment analysis effectiveness.
- Apply NLP and machine learning models, fine-tuning for accuracy.
- Implement sentiment classifiers for analysis.
- Define and use appropriate metrics to assess sentiment analysis model performance.
- Customize models to capture industry-specific language nuances in cargo and logistics.
- Implement real-time sentiment analysis for immediate insights.
- Ensure models can handle and process data in real time.
- Develop scalable solutions for handling large volumes of social media data from DHL Cargo.

- Create visualizations and reports summarizing sentiment trends over time.
- Establish a feedback loop for continuous model improvement based on user feedback and language changes.
- Work with teams to integrate sentiment analysis into processes for proactive addressing of customer concerns on social media.

Project Summary

Project: Identifying fake news using natural language processing

Role: Data Scientist

Technology: Natural Language Processing (NLP), Python, NLTK, SpaCy, TensorFlow, TF-IDF, Support Vector Machines, Random Forests, Apache Kafka, Matplotlib, BeautifulSoup.

Project Description: This project utilizes Natural Language Processing (NLP) to identify fake news by analyzing linguistic patterns, sentiment, and contextual cues in textual content. Trained on diverse datasets, the NLP models can distinguish between reliable and deceptive information in news articles and social media posts. The system enables real-time analysis for prompt detection and mitigation of misinformation, emphasizing ethical considerations to minimize biases. The goal is to provide a tool that enhances media literacy, fosters informed public discourse and mitigates the impact of fake news in the digital landscape.

Responsibilities:

- Gather diverse datasets containing both reliable and deceptive content.
- Clean and preprocess text data to ensure consistency and remove irrelevant information.
- Extract relevant features from textual content, considering linguistic patterns and contextual cues.
- Utilize techniques like TF-IDF or word embeddings for text representation.
- Develop and train Natural Language Processing (NLP) models for fake news classification.
- Choose appropriate machine learning algorithms for supervised learning tasks.
- Implement solutions for real-time analysis, incorporating streaming platforms like Apache Kafka.
- Define evaluation metrics for model performance.
- Fine-tune models based on performance feedback to enhance accuracy.
- Address biases in the machine learning models to ensure fair evaluations.
- Adhere to ethical guidelines in data processing and analysis.
- Integrate with cloud platforms (AWS, Google Cloud, Azure) for scalable computing resources and storage.
- Create visualizations to illustrate model performance and insights.
- Generate reports summarizing the outcomes of fake news identification.
- Collaborate with web scraping experts to extract relevant data from websites.
- Integrate with APIs of social media platforms or news sources for seamless data retrieval.
- Establish a feedback loop for continuous model improvement based on evolving language patterns and user feedback.
- Maintain comprehensive documentation of the entire data science process, from data collection to model deployment.
- Use Git and GitHub for version control, ensuring traceability of code changes and collaboration.

Project Summary

Project: Predictive Maintenance for SERVO MOTOR

Client: P&G

Role: Data Scientist

Technology: Python, R, Random Forest, Time-Series Analysis, SPC, Control Charts, SQL

Project Description: A servo motor's fault prediction involves collecting data from the motor via a programmable logic controller (PLC), which then undergoes data processing and statistical analysis. Machine learning algorithms are applied to predict when the motor is likely to fail, using historical and real-time data from the PLC. This prediction helps to schedule maintenance and avoid unplanned downtime of the motor.

Responsibilities:

- **Data Collection and Preprocessing:** I was responsible for collecting the relevant data related to servo motor performance and maintenance, and preprocessing the data to make it usable for analysis.
- **Data Analysis:** I was responsible for analyzing the data to identify patterns and trends that can help predict the health of the servo motor. I was using machine learning algorithms and statistical models to identify potential faults or maintenance needs before they occur.
- **Model Development:** I was responsible for developing predictive models to help detect faults or maintenance needs in the servo motor.
- **Model Deployment:** I was responsible for deploying the predictive models in production environments, integrating them with existing systems and processes.
- **Performance Monitoring:** I was responsible for monitoring the performance of the predictive models in real-time, making adjustments as needed to improve accuracy and reduce false positives.
- **Reporting:** I was responsible for creating reports and dashboards that summarize the health of the servo motor and highlight potential faults or maintenance needs.
- **Collaboration:** I was responsible for collaborating with cross-functional teams, including maintenance technicians, engineers, and managers, to ensure that the predictive models are aligned with their needs and goals.

Project: Micro-organisms Image Classification

Client: P&G(QA Lab)

Role: Data Scientist

Technology: Python, R, scikit-image, TensorFlow, Google Cloud Platform, Matplotlib

Project Description: Microorganisms Image Classification is a field of study that involves using machine learning techniques to classify microorganisms based on their images. Microorganisms include a wide variety of living organisms such as bacteria, viruses, fungi, and protozoa, which can be difficult to differentiate using traditional methods.

Responsibilities:

- **Data collection:** Collecting images of microorganisms from various sources and compiling them into a usable dataset.
- **Data preprocessing:** Preparing the dataset for analysis by cleaning, transforming, and normalizing the data.
- **Model selection:** Selecting an appropriate machine learning model for microorganism image classification.
- **Model training:** Training the selected model on the preprocessed dataset and tuning the hyperparameters for optimal performance.
- **Model evaluation:** Evaluating the performance of the trained model using appropriate metrics such as accuracy,

precision, recall, and F1-score.

- Model deployment: Deploying the trained model into a production environment and ensuring it works correctly.
- Continuous improvement: Continuously monitoring the model's performance and making necessary updates to improve its accuracy and efficiency.
- Collaboration: Collaborating with domain experts to better understand the microorganisms being classified and incorporating their knowledge into the model.
- Communication: Communicating findings and insights to stakeholders, including non-technical audiences, in a clear and concise manner.

Project: Car Number Plate Detection

Client: P&G(Transport Department)

Role: Data Scientist

Technology: Python, R, scikit-image, TensorFlow, Google Cloud Platform, YOLO, Faster R-CNN, OCR engines such as Tesseract, Matplotlib

Project Description: Car Number Plate Detection is the process of automatically detecting and extracting the number plate from an image of a car using machine learning techniques. The goal of Car Number Plate Detection is to accurately identify and extract the number plate from an image of a car, even under different lighting conditions, camera angles, and environmental factors.

Responsibilities:

- Data collection: Collecting images of cars and their number plates from various sources and compiling them into a usable dataset.
- Data preprocessing: Preparing the dataset for analysis by cleaning, transforming, and normalizing the data.
- Model selection: Selecting an appropriate machine learning model for car number plate detection.
- Model training: Training the selected model on the preprocessed dataset and tuning the hyperparameters for optimal performance.
- Model evaluation: Evaluating the performance of the trained model using appropriate metrics such as accuracy, precision, recall, and F1-score.
- Model deployment: Deploying the trained model into a production environment and ensuring it works correctly.
- Continuous improvement: Continuously monitoring the model's performance and making necessary updates to improve its accuracy and efficiency.
- Collaboration: Collaborating with domain experts to better understand the challenges of car number plate detection, including issues related to different lighting conditions, camera angles, and environmental factors.
- Communication: Communicating findings and insights to stakeholders, including non-technical audiences, clearly and concisely.

Project: Image Recognition for Packaging Inspection

Client: P&G(Tide Washing Powder Packaging)

Role: Data Scientist

Technology: Python, R, scikit-image, TensorFlow, Google Cloud Platform, YOLO, Faster R-CNN, OCR engines such as Tesseract, Matplotlib

Project Description: Image recognition for packaging inspection is a technology that uses computer vision algorithms to automatically detect and classify defects or irregularities in packaging materials. By analyzing images of packaging materials the system can identify and flag any deviations from the expected specifications, such as scratches or deformations. This

technology is widely used in the food, pharmaceutical, and manufacturing industries to ensure the quality and safety of products before they are shipped to customers. The use of image recognition for packaging inspection can help to reduce human errors, increase efficiency, and improve the overall quality control process.

Responsibilities:

- Data collection and pre-processing: Collecting relevant data and preparing it for analysis, which includes data cleaning, feature extraction, and normalization.
- Algorithm selection and development: Identifying the most appropriate algorithms for image recognition tasks and developing new algorithms if necessary.
- Model building and training: Building machine learning models for image recognition, training and fine-tuning them using appropriate data sets, and optimizing model performance.
- Deployment and testing: Deploying the trained models into production, testing their performance against benchmarks, and ensuring that they are reliable and robust.
- Collaboration and communication: Working closely with other stakeholders, such as domain experts, software developers, and project managers, to ensure that the image recognition system meets the business requirements and objectives.
- Continuous improvement: Monitoring the performance of the system, identifying opportunities for improvement, and implementing changes as necessary to enhance its accuracy and efficiency.

Project: Revenue forecasting for travel company

Client: Aspirants Travel

Role: Data Scientist

Technology: Power BI, R, SAS, decision trees, Big Data, Spark

Project Description: Revenue forecasting for a travel company involves predicting the company's future revenue based on historical sales data, market trends, and external factors. This helps the company to plan its marketing campaigns, adjust pricing strategies, and optimize operations to maximize revenue and profitability. Advanced techniques such as machine learning algorithms may be used to analyze large data sets for more accurate forecasts.

Responsibilities:

- Data collection and preprocessing: Collecting and processing data from multiple sources, such as transactional databases, social media, and external data sources, and cleaning and preparing the data for analysis.
- Statistical modeling and analysis: Using statistical modeling techniques to analyze historical sales data and identify trends and patterns that can be used to make accurate revenue forecasts.
- Machine learning modeling and analysis: Building machine learning models to predict future revenue based on historical data and external factors such as economic indicators, seasonal trends, and competitor activity.
- Data visualization and communication: Communicating findings and insights to stakeholders, such as senior management and marketing teams, through visualizations and presentations.
- Collaboration and cross-functional teamwork: Collaborating with other departments, such as marketing and sales, to gather input and insights to refine revenue forecasts and optimize revenue generation strategies.
- Continuous improvement: Monitoring the accuracy of revenue forecasts and continuously refining models to improve accuracy and efficiency.