

# SANTOSH KUMAR

## Data Scientist

8005936454

[www.github.com/santos-k](https://www.github.com/santos-k)

[@ kumarsantosh7j@gmail.com](mailto:kumarsantosh7j@gmail.com)

Gurgoan, India

[www.linkedin.com/in/santos-k](https://www.linkedin.com/in/santos-k)

[www.kaggle.com/kuchhbhi](https://www.kaggle.com/kuchhbhi)

## SUMMARY

Results-driven professional with 4.5 years of software testing expertise and 1 year in data science, specializing in machine learning, data analytics, and effective problem-solving. Committed to continuous skill acquisition and passionate about leveraging this unique skill set for a rewarding career transition in the dynamic field of data science.

## WORK EXPERIENCE

05/2020 - 11/2023

Remote

Engineer(Oct/23-Nov/23) | Engineer(Dec/22-Sep/23) | Test Engineer(May/20-Nov/21)

**Whale Cloud Technology India Privet Limited, Gurgoan**

Contributed to a government organization by applying testing skills to streamline processes and resolve operational challenges within the **telecom sector**.

- Conducted comprehensive **manual and automation testing** to ensure software quality and performance.
- Executed **database, API, web, and GUI** testing to ensure comprehensive software quality and functionality.
- Developed an automated tool to prepare test data creation which **saved 25% of testing time**.
- Achieved a **50% contribution** to the testing team's efficiency.
- Lead testing team** for nearly one year, ensuring successful project outcomes.
- Applied **Agile** methodologies for adaptable, customer-centric project delivery.
- Created detailed QA **documentation** including QA **reports**.

08/2022 - 11/2022

Remote

Data Scientist (Part-Time on Contract)

**Konnectbox, Bangalore**

Bangalore-based B2B fashion **e-commerce** startup, facilitating seamless transactions and connections.

- Scrapped** and analyzed e-commerce data for trend analysis, price comparison, and prediction.
- Conducted thorough **data cleaning**, including tag removal, handling duplicates, and addressing missing values, ensuring data quality.
- Achieved 93% accuracy in predicting product prices using a **random forest regressor** in **machine learning**.
- Developed** a deep learning-based image reverse search **recommendation tool** connected to a Flask API, providing 20 matching images for product and competitor analysis.

02/2019 - 01/2020

Gurgoan

Associate Analyst

**Wipro Limited, Gurgoan**

Applied testing skills in the **US payroll domain**, resolving Defined Contribution issues for Fortune 500 companies.

- Led diverse **software testing**, including Integration, Regression, Retesting, and UAT.
- Created and executed comprehensive **test plans**, actively engaging in internal and **client meetings**.
- Conducted **manual testing** for GUI, web, Android, and iOS applications.
- Collaborated across three teams on six projects, **earning praise** from managers and clients for **exceptional performance**.

## EDUCATION

2015 - 2019

Alwar, Rajasthan

BTech in Computer Engineering

**Rajasthan Technical University, Kota**

## SKILLS

Machine Learning

Data Analysis

Data Scrapping

Data Visualization

NLP

Python

SQL

NUMPY

PANDAS

SCIKIT-LEARN

BEAUTIFULSOAP

MATPLOTLIB

SEABORN

PLOTLY

POWER-BI

FLASK

DASH

STREAMLIT

MYSQL

MONGODB

AWS

GCP

HEROKU

JUPYTER NOTEBOOK

POSTMAN

GITHUB

## INTERNSHIPS/CERTIFICATIONS

**IOT Intern | Zeetron Networks, Jaipur**

Completed a two-month internship in IoT, gaining hands-on experience with IoT technology, devices, sensors, Arduino, and Raspberry Pi.

**Data Science Bootcamp | Udemy**

Successfully completed a comprehensive Data Science Bootcamp, acquiring essential skills in data analysis, machine learning, and statistical modeling.

**Python Certification | Amity Academy**

Attained certification in Python, covering fundamentals and key Object-Oriented Programming (OOP) concepts

## PROJECTS

### Reverse Image Search Tool for Fashion Recommendation | KonnectBox

Created a web-based reverse image search tool for fashion recommendations, facilitating the extraction and analysis of identical items from different databases. The tool accepts multiple images, CSV files with image URLs, or text-based searches, offering advanced filtering options for efficient item comparison.

- Engineered the tool to seamlessly **accept multiple images, CSV files with image URLs**, and text-based queries, providing versatile search options.
- Implemented **advanced filtering features** enabling users to refine search results based on database type, category, brand, and matching confidence level for detailed analysis.
- **Developed an intuitive and user-friendly interface** featuring card-based result displays, paginated navigation, and interactive feedback mechanisms, enhancing the user experience.
- Integrated an **API for dynamic data processing**, allowing seamless communication between the web application and the backend to fetch and display fashion recommendations.
- Implemented a **robust feedback system** where users can provide accuracy feedback on search results, enhancing the tool's learning capability and result refinement over time.
- Skills used: Web Development, Python Programming, Data Analysis, User Interface Design, API Integration.
- Tools and Technologies: Python 3.10, Dash Framework, GitHub, Pandas, JSON, Dash Bootstrap Components.

### Expense Tracker Web Application | Web Development and Visualization Project

Identified the need for a comprehensive expense tracking solution to assist users in managing financial transactions, including earnings, expenses, investments, and savings.

- Developed a powerful **web application using Flask**, incorporating user-friendly features and advanced data analysis tools.
- User-friendly **login and registration** system for secure access.
- Comprehensive **home page for recording, tracking**, and categorizing expenses, investments, and savings.
- Advanced analysis page with interactive charts for **deep insights into user spending patterns**.
- Secure session management system, **terminating sessions after 5 minutes** of inactivity for enhanced security.
- Advanced analysis page with interactive charts, including pie charts, bar graphs, and line graphs, providing deep insights into user spending patterns.
- Skills used: Web Development, Flask Framework, Python Programming, Data Visualization, Database Management, User Authentication, API Integration.
- Tools and Technologies: Python, Flask, Plotly, Pandas, HTML, CSS, JavaScript, JSON, Git, Google Cloud Platform (GCP), SQL (for database), Dash Bootstrap Components.

### Flight Price Prediction WebApp | ML Project

Developed an end-to-end solution for predicting flight booking prices, showcasing expertise in Python, data analysis, and machine learning. The project, utilizing a Kaggle dataset, features data analysis, model selection, and training, culminating in a web application for accurate flight price predictions.

- Conducted in-depth **data analysis**, employing techniques for **data cleaning, feature creation**, and exploratory data analysis.
- Created a user-friendly web application enabling users to make accurate predictions and check prediction accuracy against the test dataset.
- Explored **ML models**, selected **Random Forest Regressor**, Fine-tuned Random Forest Regressor to achieve an exceptional **92.6% accuracy**.
- Led end-to-end development of a user-friendly Flight Price Prediction WebApp.
- Successfully deployed the project on Google Cloud Platform, showcasing cloud computing proficiency.
- Skills: Machine Learning, Data Analysis and Feature Engineering, Model Optimization and Accuracy Evaluation, End-to-End Project Development.
- Tools: Python (3.6+), Scikit-Learn, Python Dash & Plotly, Google Cloud Platform (GCP), Git and GitHub, Jupyter Notebooks.

### Spam Classification WebApp | NLP Project

Developed a Spam Classification WebApp utilizing machine learning techniques, specifically MultinomialNB algorithm, achieving 95.9% accuracy and 100% precision score. Implemented comprehensive data preprocessing, exploratory data analysis, and model evaluation, showcasing effective SMS and email spam detection.

- Analyzed the SMS Spam Collection dataset, addressing **imbalances, and executed cleaning** tasks for effective preprocessing.
- Discovered patterns with **EDA, crafted features**, and visually identified distinctions between spam and non-spam messages.
- Utilized **tokenization**, special character removal, and stemming in model preparation, opting for MultinomialNB for superior performance.
- Utilized precision-focused metrics, achieving **95.9% accuracy and a perfect 100% precision score** with the MultinomialNB algorithm.
- **Skill and Tools used:** Python, Dash framework, Scikit-Learn, **NLP**, Pandas, Plotly for data analysis, Heroku and Google Cloud Platform.

## ACHIEVEMENTS



### Project Delivery

Successful project delivery, meeting objectives on time with high-quality results and positive stakeholder feedback.



### Team Player

Collaborative team player with effective communication skills, contributing to a positive and synergistic work environment.