

Contact.santosh@gmail.com<https://www.linkedin.com/in/santosh-kumar-0b766911/>

+91 9810671894

SANTOSH KUMAR

A versatile Data Scientist who delivers solutions for clients and companies.

I am an experienced **data scientist** with mathematics background and 13.10 years of experience in developing and managing IT solutions across various domains. I have 5 years of experience in practical data science with major expertise in descriptive and predictive analytics**, data mining, machine learning, deep learning and computer vision**. I have experience building fast and accurate machine learning models in Python, R and Spark. I am currently working with Nokia as solution architect for data science initiatives with prime focus of machine learning and artificial intelligence projects.

I am B.E from Karnataka university and completed a post-graduate program in Business Analytics (Data Science) from Great Lakes Institute of Management.

**SKILLS**

**Tools:** Python, R, tableau, QlikView, Knime

**Machine Learning:** Classification, Regression, Neural Networks, Clustering, CNN, RNN,Computer Vision.

**Deep Learning API:** Keras, TensorFlow, OpenCV

**NLP API:** NLTK, SpaCV

**Languages:** C#, Java, VB

**Big Data:**  Pyspark,Hive

**Statistical Methods:** Hypothesis Testing, Analysis of Variance, Principal Component Analysis

**PROFESSIONAL EXPERIENCE**

* Solution Architect(Data science), Nokia, India- June 2018 to present.
* Principal Software Engineer (Data Science), Kronos, India - July 2009 to June 2018.
* Senior Developer, Extreme It Europe, UK - Jan 2009 – May 2009.
* Senior Associate, Sapient, India- July 2008 – Nov 2008.
* Sr. Software Engineer, Open Solutions, India- April 2007 – May 2008.
* Software Engineer, R Systems, India- Nov 2004 – Dec 2006
* Analysis, exploration of large data sets in R and Python.
* Designing and implementing algorithm for performing predictions using Deep Learning, Machine Learning, computer vision and data science projects.
* Exploring new and emerging technologies in artificial Intelligence space and doing POC’s on them.
* Managing and maintain large datasets and extract meaningful insights from it
* Helping sales and pre-sales team for new initiatives.
* Mentoring, training and building the data science team.
* Collaborating with the architecture, tools, and software teams to influence the design of next-generation architectures, and programming models.
* Communicating analytical insight, recommended improvement strategies across business area.
* Business analysis for analytics projects specific to capital market and healthcare domain.

**Key Projects**

**Title-** Detect Breast Cancer using computer vision (Organization-**Kronos Solutions**)

**Description-** Design artificial intelligence system which would capable to detect breast cancer using mammographic image. Early detection of tumours is the only way to cure breast cancer thus accurate classifying reduce mortality rate brought by cancer.

**Tools-** Python, RNN, CNN, OpenCV, Keras, TensorFlow API

**Title**- Predict Daily stock market return (Organization-**Kronos Solutions**)

**Description**- This project was a POC for leading bank and objective was to design artificial intelligent system which could mimic technical analysis of stock as done by equity traders and predict the top 20 stocks giving at least 2% daily return.

**Tools-** Python, R, Deep neural network, Random Forest, Keras, TensorFlow, Yahoo Finance API, NLTK

**Title**- Predict Diabetic eye disease using deep machine learning. (Organization-**Kronos Solutions**)

**Description**- Design intelligent system that would capable of interpreting signs of diabetic in retinal photographs, potentially helping doctors screening more patients and quickly routing those who need help to a specialist.

**Tools-** Python, R, Hadoop, Deep neural network, CNN, Keras API, TensorFlow API, Hive

**Title**- Employee Retention Prediction Model (Organization-**Kronos Solutions**)

**Description**- Predict High/Low probability of attrition and time of employees on specific technologies. This help organization to plan better for future and provide support and guidelines to recruitment team to hire best fitment going forward.

**Tools-** R, Logistic Regression, Random Forest, Tableau

**Title**- HR & PAYROLL (Organization-**Kronos Solutions**)

**Description**- HR and payroll software automates administrative processes, so you can focus on managing your most valuable employee. Workforce HR™ and Workforce Payroll™ help to control labour costs, minimize compliance risk, and manage talent with ease

**Tools-**, R, C#, Vb.net, Crystal report, Exago BI reporting, sql server, Asp.net, Angularjs.

**Title-** Finance Management and Task Scheduler Service (Organization-**Extreme-It**)

**Description-** It automates Finance, resourcing, time sheet, expenses. Task scheduler is also part of this tool that will schedule the services to send the notification mail and send the reminder to concern person.

**Tools-** C#, Vb.net, Crystal report, sql server, Asp.net

**Title-** Profit Vision (Organization-**Open Solutions**)

**Description**- This application helps bank and Credit Union to get the details for profitable customer and help to find the profitable product for bank. Application works with core bank data and import all the required database.

**Tools-** Vb.net, Active Report, sql server, Asp.net

**Title-** Billing utility (Organization-**R Systems**)

**Description-** This module allows with their business practices, from initial contact with a citizen or business, to the financial accounting for each transaction. This module consists of Sub modules viz. Accounts, Bills, Payments, Transactions, Delinquency, Deposit & Interests.

**EDUCATION**

Great Lakes | **PG in Business Analytics** (**Data science)**

KARNATAKA University | **B.E in Instrumentation Technology**

**ACADEMIC PROJECTS (DATA SCIENCE)**

* **Credit Scorecard Model:** Built a credit risk model on dataset of Indian companies to predict the probability of default using logistic regression framework. Information package in R was used to perform variable screening and exploratory analysis with WOE and IV. Stepwise Logistic regression model technique was used to build the model by first building the model using all the important variables and then removing most insignificant variable at each step.
* **Text Analysis of Twitter Data:** Twitter analysis and text mining of tweets data for a famous brand by extracting 1000 latest tweets using R and Python. Python’s ‘Tweepy’ package was used to download tweets data from twitter and R’s’ tm’ and ‘qdap’ package was used for cleaning and analysis of tweets data. Various analysis like – frequent term analysis, topic modelling, word association and sentiment analysis were performed.
* **Next Best Products for existing customer:** Design a decision engine that guides each inbound and outbound customer action and communication for every channel and line of business. transform approach from a product-centric focus to a more customer-centric view. It means product offerings are tailored for each customer.
* **Election Prediction Model:** Built a model to predict the outcome of 2014 Lok Sabha election and identify the factors impacting the chances of candidates winning an election. Data for this project was collected from various sources like election commission website, census website, My Neta website etc. A logistic regression model was built on this dataset and the probability of a candidate winning election in a constituency is determined.