

HARISH KUMAWAT

Lead Data Scientist at Msys Technologies

Email: harishk9366@gmail.com

Mobile: +91-7014934341

Notice Period: 30 Days

Github ID: [harishkumawat2610](#)

LinkedIn: [harishkumawat](#)

YouTube: [YouTube channel](#)

Professional Summary

- ✓ Have **5 years** of IT industry experience as **Data Scientist**.
- ✓ Good knowledge in **machine learning algorithms** (recommendation system, Classification, Regression, Object Detection) , **Deep learning** and **Computer Vision**.
- ✓ Knowledge on **Heroku, MongoDB, AWS Lambda** and deploying ML models.
- ✓ Good Understanding of **predictive modeling** and Machine learning algorithm.

Roles and Responsibilities

- ✓ Understanding the problem statement for the Machine learning and deep learning Models.
- ✓ Getting the data from client and validate using **feature engineering**.
- ✓ Use exploratory **data analysis** on the data, finding the appropriate algorithm for the model.
- ✓ Validate the model and presentation to client.
- ✓ Good understanding of **data mining, cleaning, modeling**.
- ✓ Worked on Python to **scrape, clean** and analysis large datasets.

Technical skills

Domain	:	Machine learning (Linear regression , Logistic Regression, Clustering, Random forest and decision tree , KNN) , Ensemble learning (Bagging, Boosting) , NLP and Deep learning (ANN,RNN,CNN,LSTM,Encoder,Decoder,Autoencoder), Object detection techniques (Gradient descent and Stochastic Gradient Descent), OpenCV, YoLo, Transfer learning (ResNet-50), Tkinter, Scikit-learn, Pandas, Numpy, Computer Vision , Image processing, Matplotlib, Flask, Pyautogui, Selenium, BeautifulSoup4, Nltk, Spacy, word2vec, texthero, Crawler, Numba, YoLo-annotation Character-level Embedding, Streamlit, AWS Textract, Aws Lambda, Transformers Pipeline, Plotly and familiar with Cyber Security .
Platforms	:	Linux and Windows.
Frameworks	:	TensorFlow, NLTK.
Tools	:	Pycharm, VS Code, Jupyter Notebook, Google Colab, Anaconda and Spyder.
Database	:	MySQL and familiar with MongoDB and Cassandra
Language	:	English, Hindi, Rajasthani

Educational Qualification

B.Tech (Bachelor of Technology) in Computer Science and Engineering (2015-2019)

Government Engineering College Ajmer, Rajasthan

Senior Secondary (12th PCM 2013-2014)

Noble Academy Sr. Sec. School, Dantaramgarh, Sikar (Rajasthan)

Msys Technologies, Bangalore (1st April 2022 – present)

- I am working as a **Lead Data Scientist** in Msys Technologies.
- **Worms Connect-ML** : Experience on **AWS Textract API** and **AWS Lambda** for Extract Required Fields from **invoice pdfs** and Worked on **S3 buckets**, **AWS RDS** (Relational Database Service) and Analysis **CloudWatch** .
- **Feedback NLP Model** : Worked on **Sentiment Analysis** (Sentiment detection) and **Department Detection** with **Dashboard** (**Plotly graphs**) and worked on **Transformer Pipeline** library for **sentiment analysis**.

SequireTek It Solution Pvt. Ltd, Bangalore: - (15th march 2021 – 31th 2022)

- **Web Model**: Experience on **Character-Level CNN** (Character-Level Embedding) model for **Attack Detection** (SQL Injection, XSS (Cross-site Scripting), DT (Directory Traversal), FI (File Inclusion) and Clean. Worked on **1 lakh** labeled **request URL dataset** for model.
- **Unified Model**: - Worked on **LightGBM** model for Unified project with approximate **600-700 labels** for unified use cases. Generated **synthetics data sheet** (mapping) according to conditions and trained LightGBM model on synthetics data sheet.
- **Port Scanning Script**: - Worked on python script for **port scanning detection** and also gathering data according to script condition for ML model.
- **Critical Port Scanning Script**:- Worked on python script for Critical port scanning detection and also **gathering data** according to script condition for **ML model**.

GlobusSoft Technologies, Bangalore (9 Jan 2020 – 6 march 2021)

- **Created a Reverse image Search Engine for PowerAdspy**:- Worked on **ResNet-50** (pre-trained Convolution neural network) model for **transfer learning** and **extract the features**, and then use a metric to calculate the error rate like **Euclidean distance**. Trained **nearest-neighbors model (KNN)** using the **brute-force algorithms** to find the nearest five neighbors based on **Euclidean distance**. Performed **feature extraction and similarity** search on **Caltech101** (approximate 9,000 images) and **Caltech256** datasets (approximately 30,000 images). [Demo video](#)
- **Created Ads Scraper from YouTube and Quora**: - Worked on **Selenium** and **BeautifulSoup4** for python Script.
- **Created own OCR**: - **Tesseract** (Open Source) is good for **black & white** and **gray** images but not good for **Colorful images** (different fonts and size). Created Crawler for **Data Gathering** using Selenium and BeautifulSoup4. Experience on Yolo for my own OCR and added **YoLo-annotation** of Alphabet letter (A to Z, a-z and 0-9) and added customs objects in **YoLo**.
- Created Crawlers for **Data Gathering (Selenium and BeautifulSoup4)**.

Adhoc Networks Infotech Private Limited, Jaipur (Data Scientist Intern) (Aug 2018 – Nov 2019)

- Created an **Automatic Attendance System** with **Tkinter** (Tkinter used for Python GUI). [Demo](#)
- Created **MySQL Database** for Automatic System with **Website** interface. [Demo](#)
- Improved **Automatic Attendance System** with **CCTV Camera**. [Demo](#)
- Created a **NLP chat-bot** for Adhoc Networks Infotech Private Limited. [Demo](#)
- Created a **Car Price Prediction model** and Deployed on **Heroku** platform. [Demo](#)