



Immense interest in technologies.
Enthusiast in the field of Data Science,
Machine Learning and Analytics.



harpal.kalsi@gmail.com



9866 26 88 91



<https://www.linkedin.com/in/harpal-singh-kalsi-9447751a/>

CERTIFICATIONS

- ❖ Machine Learning Certificate from **Stanford University** (Offered through Coursera)
- ❖ Project Management Professional (PMP) certified by **Project Management Institute (PMI), USA**

PROFESSIONAL SUMMARY

- ❖ Over 4 years (out of around 13 years of experience) in Data Pre-processing, Data Science, Machine Learning models
- ❖ In-depth understanding of Data Science, Statistical Data models and Machine learning
- ❖ Supervised and Unsupervised Machine learning algorithms and techniques
- ❖ Programming languages Python and R with real-world customer projects
- ❖ PGP in Machine Learning, Analytics & Optimization from INSOFE
- ❖ Basic understanding of Kafka and Flume
- ❖ Valid B1 Visa

HARPAL SINGH KALSI

DATA SCIENTIST

WORK EXPERIENCE



EMIDS TECHNOLOGY, Data Scientist / PSE

Jan 2019 - Present

Project: Breast Cancer Prediction with Genome data

Objective: To predict the presence of breast cancer in a patient.
This project was delivered as Python package.

Role: As a Data scientist:

- Feature engineering
- Sparse representation of data
- Model building

Tools / Techniques: Python, Logistic Regression, Decision Trees, Random Forest, Sparse vectorization, SciPy

Project: Symptoms Analyzer in a call recording (in .WAV) file

Objective: To find out the symptoms which are spoken in the call from a patient to the nursing facility.
This project was delivered as Python package.

Role: As a Data scientist:

- Converting .WAV file to text
- Preprocessing / cleaning of text
- Model building

Tools / Techniques: Python, NLTK, DBOW, Logistic Regression

SKILLS:

Supervised Machine Learning:

- ❖ Linear Regression
- ❖ Logistic Regression
- ❖ SVM
- ❖ Decision Trees
- ❖ Random Forest
- ❖ Naïve Bayes
- ❖ XGBoost GBM
- ❖ Neural Networks (NN)
- ❖ Deep Learning
- ❖ Convolution NN

Unsupervised Machine Learning:

- ❖ K-Means
- ❖ Hierarchical

Big Data Technology:

- ❖ Hadoop
- ❖ Spark
- ❖ MapReduce
- ❖ HDFS
- ❖ YARN
- ❖ Pig
- ❖ Hive
- ❖ Sqoop

Languages:

- ❖ R
- ❖ Python

Tools:

- ❖ Tableau
- ❖ Qlik Sense
- ❖ Hortonworks
- ❖ Putty
- ❖ WinSCP
- ❖ Confluence
- ❖ MS Office

Project: Auto-scaling of containers

Objective: To build a ML model which can predict the time taken to process the claims by the server and accordingly set the number of containers needed from any point of time onwards, thereby implementing auto-scaling.

Role: As a Data scientist:

- Reading data from Kafka message queue
- Cleaning and preprocessing of data
- Model Building, model tuning

Tools / Techniques: Python, Linear Regression, Random Forest

Project: Census Forecasting for optimal staffing

Objective: The number of patients that are in a hospital, or 'census', is difficult to predict, but extremely important when determining how to staff a hospital or department in the next day, days, or week. Data available often varies significantly from practice to practice, so a single solution is often difficult. The objective of this project is to forecast patient load based on time series analysis with minimal information.

Role: As a Data scientist:

- Model Building,
- Model tuning

Tools / Techniques: Python, Monte Carlo

NIIT TECHNOLOGIES, Data Scientist, Architect

Jul 2018 – Dec 2018

Project: Object Recognition using Deep Learning for a UK based Insurance client

Objective: To first classify the image into Roof and Kitchen. Further classify roof into good, damaged and tarred roof. Kitchen into kitchen with and without appliances. The project then automatically identifies and calculates the damaged area. This was an end-to-end project implemented on HTML and deployed using Django framework on AWS. The IP of AWS has been made public which can be accessed from anywhere and everywhere (including mobile as well). This was done in Python using Deep Learning (CNN) using Keras and CV2 algorithm.

Role: As a Data scientist:

- Reading images, pre-processing, data augmentation
- Model Building, model tuning
- Model deployment using Django framework

Tools / Techniques: Python, CV2, CNN, Django, HTML

Functional Skill set:

- ❖ Excellent analytical and problem-solving skills.
- ❖ Ability and zeal to learn new technologies.
- ❖ Self-motivated and quick learner.
- ❖ Excellent communication skills



STRENGTHS:

- ❖ Analytical
- ❖ Communication
- ❖ Team player
- ❖ Flexible

ORACLE INDIA PVT. LTD., Data Scientist / Manager

Oct 2014 – June 2018

Project: Oracle's Primavera Unifier is the best-in-class project. Lifecycle management solution for capital planning, project delivery, cost control, and facilities and real estate management. It provides governance across all project phases, from planning, building to operations and maintenance.

Role: As a Data scientist:

- Data cleaning / pre-processing in R / Python
- Supervised Machine learning model building in R / Python
- Acting as Data Scientist POC
- Worked on several customer Data Science projects involving Linear Regression, Logistic regression etc.

Project: Sentiment analysis for an online retail company

Objective: To do sentiment analysis of twitter tweets for the company and classify it as Positive, Neutral and Negative. Help the client understand and take corrective and preventive steps.

Role: As a Data scientist:

- Text Cleaning / pre-processing in Python
- Model Building

Tools / Techniques: Python, NLTK, Naïve Bayes

Project: Analyzing click stream data

Objective: To understand click stream data on the client website and classify the behavior into human being and non-human being (Bot).

Tools / Techniques: Python, classification techniques

Project: IOT Signals Classification

Objective: To classify the incoming IOT signals to various classes (Database, Hardware, Network Layer Issue), which in-turn will help the customer to reduce the human dependency and minimize the TAT.

Tools / Techniques: Logistic Regression, SVM and Random Forest

Project: Indian Telecom Company

Objective: To predict customer churn

Tools / Techniques: Logistic Regression, Random Forest, GBM



AWARDS & ACHIEVEMENTS

- ❖ Facial Recognition using ML: **Winner**, Hackathon, emids technology
- ❖ Honored with the most prestigious and highly visible award, “**Living Our Values**” from **United Health Group** for demonstrating outstanding integrity and excellent project management skills
- ❖ Secured **2nd position** in MCA, whole Kanpur University
- ❖ College **topper** in BCA

Project: US Motor Manufacturing Company

Objective: To predict the price of the used cars.

Tools / Techniques: Linear Regression, Decision Trees

Project: Various clients

Objective: To do performance predictions of applying the patch in customer environment (on cloud) helping customer to estimate the downtime time.

Tools / Techniques: Linear regression



UNITEDHEALTH GROUP, Project Manager

Mar 2010 – Oct 2014



Oracle INDIA PVT. LTD., QA Engineer

Jan 2007 – Mar 2010

EDUCATION

MS (Software Systems)

BITS PILANI, Rajasthan

MCA

Kanpur University (Chatrapati Sahu Ji Maharaj University)