# My Portfolio

Table of Contents

1. Introduction**1**
2. My Projects**4**
   1. Natural Language Processing 6
      1. Emotion and Personality Trait Detection from social media text6
      2. Automatic Ticket Classification for a Banking Solution 6
      3. Disease-Treatment Matching for HealthCare application6
   2. Computer Vision:
      1. Skin Cancer Detection6
      2. Gesture Recognition6
   3. Other projects:
      1. Customer churn classifier for the Telecom Industry6
      2. Real-estate price prediction and analysis6
      3. Bike-sharing demand prediction and analysis6

Type chapter title (level 2)5

Type chapter title (level 3)6

1. **Introduction:**

I am a software developer, based out of Bangalore, India, with more than 13 years of experience in designing and developing software features and products that meet or exceed customer expectations.

I am currently developing AI/ML solutions to help customers with their business requirements, with a key focus on natural language processing (NLP), deep learning, computer vision, and linear and logistic regression.

I hold a post-graduate Master of Science (M.Sc.) degree in Machine Learning and Artificial Intelligence (ML and AI) from Liverpool John Moores University (LJMU), United Kingdom and a graduate engineering degree, B.E. in Computer Science and Engineering, from Anna University, Chennai, India.

My key skills include:

* Python.
* Java.
* Natural language processing.
* Deep learning.

I look forward to collaborating with you to help you with fulfil your business needs in a timely manner.

1. **My projects:**
   1. **Natural Language Processing:**
      1. **Emotion and Personality Trait Detection from social media text:**

* Development of a hybrid deep-learning classification model to address existing research gaps and build a better classifier to determine the personality trait or emotion being conveyed by input lines of text gathered from social media..
* Programming language used: Python.
* Python libraries used: NLTK, Keras, Gensim (FastText, Word2Vec), GloVe, Sklearn etc.
* Datasets used:
  + MBTI
  + TEC
  + ISEAR
* Processing steps:
  + Data pre-processing steps (text cleanup, lemmatization, word tokenization etc).
  + Data splitting (train-validation-test split).
  + Text tokenization.
  + Label encoding.
  + Class imbalance handling.
  + Word embedding (FastText, GloVe, Word2Vec).
  + CNN layers.
  + Bi-directional LSTM (or) Bi-directional GRU layer.
  + Classification layer (softmax).
  + Model performance evaluation (precision, recall, accuracy, F1 score, AUROC).
    1. **Automatic Ticket Classification for a Banking Solution:**
* Development and evaluation of classifier models, to classify customer complaints based on the products (or) services, by first using NLP techniques such as topic modelling and NMF, to identify the topics (or) classes to segregate the data based on, and then building multiple supervised ML classifiers model using the same, and finally identifying the best performing classifier.
* Programming language used: Python.
* Python libraries used: Sklearn, Pandas, Wordcloud, Pandas etc.
* Dataset used: dataset of complaints received by banking customers.
* Processing steps:
  + Data loading and text pre-processing.
  + Exploratory data analysis (EDA).
  + Feature extraction and topic modelling.
  + Model building (supervised), training and evaluation.
  + Model inference.
    1. **Disease-Treatment Matching for HealthCare application:**
* A classifier model to determine the correct treatment for the input disease symptoms was developed, using the Conditional Random Field (CRF) technique in NLP. The output was a dictionary of the symptoms and their corresponding medical treatment.
* Programming language used: Python.
* Python libraries used: sklearn\_crfsuite, spacy etc.
* Processing steps:
  + Data pre-processing.
  + Concept identification using PoS tagging.
  + Feature extraction, model building and evaluation using CRF.
  + Disease-Treatment dictionary creation.
  1. **Computer Vision:**
     1. **Skin Cancer Detection:**
        + Built a deep-learning based classifier to correctly detect the type of skin cancer from an input image.
        + Programming language: Python.
        + Python libraries used: tensorflow, keras, numpy, augmentor etc.
        + Processing steps:
  + Data splitting.
  + Handling class imbalance using the augmentor library.
  + Improving the training data quality through image rotation, flipping etc.
  + Model training and evaluation.
    1. **Gesture Recognition:**
       - Developed a machine learning model (deep learning model) to correctly identify the gesture being performed, from an input set of images that represent a video.
       - Programming language used: Python.
       - Python libraries used: keras, tensorflow, cv2 etc.
       - Processing steps:
  + Data augmentation and splitting.
  + Model training and evaluation.
  1. **Other projects:**
     1. **Customer churn classifier for the Telecom Industry:**
        + Built a prediction model for a leading company in the telecom industry to predict churn of high value customers.
        + Programming language used: Python.
        + Python libraries used: sklearn, matplotlib etc.
        + Processing steps:
  + Initial analysis, data visualization and data cleanup.
  + Building different models with principal component analysis (PCA): logistic regression, random forest etc.
  + Building different models without principal component analysis (PCA): logistic regression, random forest etc.
  + Models’ evaluation and selection of the best performing model.
  + Important churn factors’ identification.
    1. **Real-estate price prediction and analysis:**
       - Development of an ML model to better predict housing sale prices for the real-estate industry and to better identify the key factors affecting the same.
       - Programming language used: Python.
       - Python libraries used: sklearn, matplotlib etc.
       - Model processing steps:
  + Data cleanup, visualization, train and test data creation.
  + Regression model building using Ridge and Lasso, and evaluation.
    1. **Bike-sharing demand prediction and analysis:**
       - Built a predictive model, to better understand the driving factors behind the demand for shared bikes and enabling the bike sharing provider to formulate a better business strategy to meet the customer’s expectations.
       - Programming language used: Python.
       - Python libraries used: sklearn, matplotlib, numpy etc.
       - Processing steps:
         * Data pre-processing: visualization, cleanup etc.
         * Train-test data creation.
         * Linear model building, with recursive feature elimination.
         * Model evaluation and further refinement to obtain the best performing model.
         * Inference from the resulting model.