SURINDER KUMAR

Ph. no: +91 7030534712
E-mail: imsurinder90@gmail.com
GitHub : https://github.com/imsurinder90
LinkedIn: https://www.linkedin.com/in/surinderkumar1990/

Address: H.No. – 24,
Bhainsa Tibba,
Panchkula – 134114

PROFILE SUMMARY

Highly motivated and enthusiastic Lead Data Scientist with 7+ years of work experience in resolving complex problem In multiple domains by taking own initiatives by understanding Client's Business problems and requirements.

EDUCATION

B.TECH (CSE) from Punjab Technical University securing 80.4% (April, 2012)

12th from C.B.S.E securing 64% (July, 2008)

10th from C.B.S.E securing 66.5% (March, 2006)

OPENSOURCE CONTRIBUTIONS

- ✓ Contributed to **RASA NLU** and **RASA Core** Chat-bot Framework like DialogFlow. Appreciated with RASA **Pin** and RASA **Stickers** for the contributions.
- ✓ In Top 3 contributors of **pgAdmin4**, a GUI tool for **PostgreSQL** database.

PROFESSIONAL EXPERIENCE

ECLERX Services, Chandigarh, India   **Process Manager, Feb 2018 – Oct, 2019**

Job Responsibilities: Predictive analytics and Chat bot projects which built on Natural language processing, Natural language understanding techniques. Handles and **mentor** the team of **size 5**.

Projects Summary:

- ✓ **Predictive Models** in retail banking and financial domains: Used algorithms such as Logistic Regression, Decision tree, Random forests, SVM and Artificial neural networks to build Classification models and deployment.
- ✓ **Image classification** in healthcare domain: Used Convolutional Neural Networks to classify the category of images.
- ✓ **Semantic Role Labeling:** Built models for named entity recognition, data tagging, extracting time phrases, KPI and Roll up region entities. Taken care of synonyms using the lookup tables.
- ✓ **Voice Remote**: Automated TV, Projector & Polycom remotes in Boardroom using **Alexa Echodot** and **Raspberry Pi**.
- ✓ **Chatter Bot:** Built a chat bot which acts like a virtual customer to train the agents/employees (voice and chat teams) with questions and answering using Classification techniques such as SVM, Random Forest and a mix of keyword matching, Fuzzy string matching & text similarity algorithms.
- ✓ **Recruitment Bot on NLU**: Built an interactive Chabot from ground up using Rasa NLU library which is identical for extracting intent and entities from the input data.
- ✓ **Named Entity Recognizer**: Developed NER using Deep Learning based **LSTM** to identify Skills, Roles, Company, Notice Period, Locations patterns from the parsed resume. Extracting useful information from Resume and updating the job seeker profile. **LSTM Bidirectional model** is used to extract entities from the text data. **Conditional Random Fields** (CRFs) used to extract the entities with much greater accuracy.

ENTERPRISEDB, Pune, India   **Software Engineer, June 2015 – Jan 2018**

Job Responsibilities: Predictive modeling and Anomaly detection which constitutes data preparation, data transformation, exploratory data analysis, data cleaning/preprocessing, feature engineering, model building, model performance evaluation, parameter tuning, model selection, interpretation of model, model visualization and creating result reports using tools like Seaborn, Matplotlib & MS Excel etc.

Projects Summary:

- ✓ **Predictive Modelling**: Involved in a team of size 4 and developed various predictive models to predict the CPU and Network usage based on CPU metrics, usage metrics of historical data.
- ✓ Classify If new Customer will subscribe a plan or not. Historical data of existing customer is used. Techniques used such as SVM, Decision Trees, Logistic Regression, Naïve Bayes, Bagging and Boosting.
- ✓ **Customer Segmentation**: Used K-Means algorithm to create groups of customers to implement marketing plans, product pricing, promotions. Other techniques used LDA, LSA etc.

- ✓ **Anomaly (Intrusion) Detection**: Patterns are observed for disk usage, memory usage, database size growth. It tracks these features over a period of time and if there are changes or fluctuations in standard deviation values, means there is something wrong and an immediate action is required. This alert is sent to the administrator. Techniques such as Clustering, one Support Vector Machine, and K-nearest Neighbor is used.

Avant Garde Services, Mohali, India   **Web Developer, Dec 2013 - May 2015**

Job Responsibilities: Predictive analytics projects which constitutes exploratory data analysis, data cleaning/preprocessing, feature engineering, model building, model performance evaluation, parameter tuning, and model selection, interpretation of model, visualization and classification reports.

Projects Summary:

- ✓ **Churn Prediction**: Tree based machine learning models are used. Handled large imbalanced data.
- ✓ **Sentiment analysis** to identify the user's tone and co-reference resolution to analyze the large data.
- ✓ **Feature Extraction**: Extracted features (word parts, simplified POS tags, lower/title/upper flags, features of nearby words) and convert them to sklearn-crfsuite format - each sentence should be converted to a list of dicts.
- ✓ Text classification: Used NLTK for text cleaning, used techniques such as Bag of words, LSTM, Word Embedding, Word2Vec, re libraries & Naïve Bayes.

Solutions Beyond, Mohali, India   **Software Developer, Jan 2012 - Nov 2013**

Job Responsibilities: Python and Data scraping based projects.

Projects Summary:

- ✓ Scraped data using beautiful Soup library and regex from the Google AdWords tool for SEO, Page ranking product and stored into the CSV and databased.
- ✓ Scraped amazon reviews using Scrapy library for client project.
- ✓ Build a client website using Flask with user authentication, created modules and packages.

PROFESSIONAL SKILLS

- ✓ Strong knowledge in building Statistical/Predictive models using Supervised and Unsupervised Machine learning algorithms: Regression, Classification, Clustering and Association Rule Mining.
- ✓ Adept in Deep Learning methods such as artificial neural networks, Convolutional neural networks.
- ✓ Good knowledge in Natural Language Processing – Text classification, Sentiment/opinion mining, Topic modeling (Text mining), POS tagging, Named entity recognition, Semantic role labeling and Co-reference resolution and Natural Language Understanding, Image processing.
- ✓ Libraries used: **Scikit-learn**, numpy, **pandas**, nltk, **Tensorflow**, **keras**, **spacy**, rasa_nlu, rasa_core, chatterbot, scipy, **practalnlptools**, matplotlib, re etc..
- ✓ Model Optimization: Model performance evaluation (K-Fold Cross-validation), metrics evaluation, Parameter tuning (Grid search), Regularization, Dimensionality reduction (PCA, LDA).
- ✓ Exploratory Data Analysis: univariate, bivariate and multivariate analysis.
- ✓ Data Visualization: scatter plots, box plots, histograms, bar charts, graphs etc.

TRAININGS AND CERTIFICATIONS

- ✓ Completed courses on **Machine Learning & Deep Learning** from **Coursera**.
- ✓ Completed courses on **Statistical Analysis in R** from **Stanford University**.
- ✓ Certified in **PostgreSQL** Database (Jun 2017 – Present & License EDB24930)

REFERENCES: Available upon Request