**SURINDER KUMAR**

|  |  |
| --- | --- |
| Ph. no: +91 7030534712 | Address: #24, MDC |
| E-mail: [imsurinder90@gmail.com](mailto:imsurinder90@gmail.com) | Sector-5, Manimajra |
| GitHub : <https://github.com/imsurinder90> | Panchkula - 134107 |

# 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 (COMPUTER SCIENCE AND ENGINEERING), Apr 2012, 80.4%**

Punjab Technical University, PUNJAB, INDIA

**PROFESSIONAL EXPERIENCE**

### ECLERX SERVICES, CHANDIGARH, INDIA

**LEAD DATA SCIENTIST, Feb 2018 – till date**

**JOB Responsibilities:** Text analytics and Chat bot projects which built on Natural language processing, Natural language understanding techniques.

**Programming Tools used:** Python, PyCharm, Vim

**Projects Summary:**

* **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.
* Sentiment analysis to identify the user’s tone and co-reference resolution to analyze the large data.
* And also built word clouds using techniques such as tf-idf and bi-gram, n-grams to analyze text and find out the NPS detractors for customers. Deployment using Apache with wsgi service.
* **Chatter Bot:** Built a chat bot which acts like a virtual customer to train the agents/employees (voice and chat teams) with questions and answers using Classification techniques such as SVM, Random Forest and a mix of keyword matching, Fuzzy string matching & text similarity algorithms.
* **Recruitment Bot on Natural Language Understanding**: Built an interactive Chabot from the scratch using Rasa NLU library which is identical for extracting intent and entities from the input data. The bot was intended to make easier the user’s interaction for job search and apply the listed jobs based on user information.
* **Named Entity Recognizer**: Developed NER using Deep Learning based **LSTM** to identify Skills, Roles, Company, Notice Period, Locations patterns from the parsed resume. Also extracting the useful information from Resume and updating the job seeker profile. **LSTM Bidirectional model** is used to extract the entities from the text data. For Training NER model user data is used with tagged skills, roles and other entities. **Conditional Random Fields**(CRFs) used to extract the entities with much greater accuracy.
* **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.

**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.

**Programming Tools used:** Python, SQL server, Advanced Excel

**Projects Summary:**

* **Predictive Modeling**

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.
* **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.
* **Churn Prediction**: Tree based machine learning models are used. Handled large imbalanced data.

### Avant Garde Services, CHANDIGARH, INDIA Software 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, model selection, interpretation of model, visualization and classification reports.

**Programming Tools used:** Python

### 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.
* Image classification in healthcare domain: Used Convolutional Neural Networks to classify the category of images.
* Clustering/Segmentation: Used K-Means algorithm to create groups of customers to implement marketing plans, product pricing, promotions. Other techniques used LDA, LSA etc.
* Text classification: Used NLTK for text cleaning, used techniques such as Bag of words, LSTM, Word Embedding, Word2Vec, re libraries & Naïve Bayes.

### Solutions Beyond, CHANDIGARH, INDIA Software Developer, Jan 2012—May 2013

**JOB Responsibilities:** Python and Data scraping based projects.

**Programming Tools used:** Python, Scrapy, BeautifulSoup, Regex, Flask

### 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

* Good knowledge in Descriptive Statistics- measures of central tendency & measures of dispersion, Inferential Statistics.
* 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.

**TECHNICAL TOOLS DATABASE TOOLS VISUALIZATION & REPORTING**

|  |  |  |
| --- | --- | --- |
| * Python and R | * SQL, PostgreSQL | * Matplotlib |
| * MS Office suite |  | * Seaborn |
| * OS: Windows |  |  |

## 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)
* Utilized other resources such as YouTube channels, forums, blogs etc. to gain knowledge.

## REFERENCES: Available upon Request