**Sarang Sanjay Gujar**

(682) 223-3474 | [gujar.sarang@gmail.com](mailto:gujar.sarang@gmail.com)

**Professional Summary:**

* **Data Science Professional** with a **5+ years** of proven record in solving real-world business challenges using data analytics and advanced data-driven methodologies. Demonstrated expertise in deploying complex machine learning and statistical modelling algorithms for extracting valuable insights for key stakeholders and organizational leadership along with proficiency in **Natural Language Processing**, **Pattern Recognition**, **Robotic Process Automation**, **Time Series Analysis and Forecasting**.
* Experience in Descriptive Analysis Problems like Frequent **Pattern Mining, Clustering, Outlier Detection.**
* Experience in leveraging machine learning models such as **Supervised learning (linear and logistic regression, boosted decision trees, Support Vector Machines, neural networks), Unsupervised learning (clustering, dimensionality reduction, recommender systems).**
* Worked with several python packages such as **NumPy**, **SciPy**, **Matplotlib**, **Seaborn**, **Beautiful Soup**, **Urllib2**, **Requests**, **Pandas**, **NLTK**, **Scikit-Learn**, **Keras**, **TensorFlow** and **CSV**.
* Experienced in developing web-based application using **Python2/3** with **Django**, **Flask**, **Pyramid**, **Bottle**.
* Experienced in **Full stack SDLC**, **Architecting scalable platforms** and **Object-Oriented designing and programming.**
* Hands on expertise on various **Design Patterns**, **Multi-threading**, **Collections** and **UML** concepts in **Python** and **Java**.
* Experienced in writing **Spark Applications** in Python using **PySpark**.
* Proficient using **Python** for **statistical analysis** and **visualisation** in **Tableau** and **Qlikview** to create visually powerful and actionable interactive reports and dashboards.
* Experience in implementing Model View Control (MVC) architecture using Server-side application like **Django**, **Flask**, **Bottle** and **Pyramid** for developing web applications.
* Worked on automation using the **Python as a scripting language**, **Git on Cygwin32** and **XML**.
* Expertise in **UNIX** and **Linux Shell** environments using **command line utilities**.
* Proficient in using editors **Eclipse**, **PyCharm**, **Jupyter**, **Spyder**, **PyScripter**, **Sublime Text** and **Notepad++**.
* Experience in version control systems using **GIT**, **GITHUB** and **SVN**.
* Experience in working with **Relational DB (RDBMS)** such as **SQLite**, **MYSQL**, **Workbench**, **PostgreSQL** and **No-SQL** database like **MongoDB**, **DynamoDB,** **CassandraDB, CosmoDB.**
* Professional in designing and maintaining complex **SQL queries and developing PL/SQL stored procedures.**
* Good experience with cloud platforms such as **AWS, Azure** and **Google Cloud Platform**.
* Experience with **WAMP** and **LAMP** environments.
* Worked on **Jira** for bug tracking in a test driven development.
* Experience in working with different environments like **Windows**, **MacOs**, **Linux**, **UNIX**.
* Strong analytical skills accompanied by good written and interpersonal communication, documentation and testing skills.

**Technical Skills:**

|  |  |
| --- | --- |
| Languages: | Python, R, C++, Java, Kotlin, SQL, JavaScript. |
| Machine Learning: | Logistic Regression, Linear Regression, Decision Tree, Naïve Bayes, SVM, Neural Networks, Sentiment Analysis, K-Means Clustering and KNN, NLP, Time Series Analysis and Forecasting. |
| Statistical Methods: | Descriptive statistics, Hypothesis Testing, Chi-square testing, F-tests, Confidence Intervals, Bayes Law, Dimensionality Reduction, ROC curve |
| Frameworks: | Django, Flask, Pyramid, PyJamas. |
| Python Packages: | NumPy, SciPy, Scikit-learn, NLTK, TensorFlow, Keras,SASPy, Statsmodels, Seaborn, Matplotlib, Urllib2, Requests, Pandas, Beautiful Soup, CSV, MySQL Connector, PyMongo, PyDynamo, and Python-twitter. |
| Cloud Platforms: | AWS, Azure, Google Cloud Platform, IBM Cloud (Watson). |
| Database: | MySQL, PostgreSQL, SQLite, Mongo DB, Dynamo DB, Cassandra DB, Cosmos DB. |
| IDEs: | Anaconda, Jupyter, Spyder, PyCharm, Eclipse, Webstrom, Sublime Text. |
| Version Control: | GIT, Github and SVN |
| Web Server: | Apache Tomcat, Web logic, AWS Instance, Azure Instance, LAMP, WAMP. |
| Testing Tools: | Jira, Bugzilla, Jasmine. |
| Operating Systems: | Windows, MacOs, Linux and Unix. |

**Career Summary:**

**Client: Becton Dickinson and Company, Franklin Lakes, NJ December 2018 – Present**

**Role: Machine Learning Engineer /Data Scientist**

**Description**: Becton, Dickinson and Company is an American medical technology company that manufactures and sells medical devices, instrument systems, and reagents. Where I was working with the Data science team for improving and scaling business operations within the organization globally by developing state-of-the-art machine learning solutions along with gaining Insights from the organizations data to elevate performance and overall functionality.

**Responsibilities**:

* Responsible for developing end-to-end metrics system for data analytics and data automation with customized visualization using Python, Automation Anywhere and Power BI.
* Implemented methods of NLP for sentimental analysis, dependency parsing, entity recognition, lemmatization, stemming, tokenizing and POS tagging using spaCy, Gensim and NLTK for understanding customer queries/interactions with the products.
* Implemented various flavors of machine learning algorithms such as (I) supervised learning (linear and logistic regression, decision trees, support vector machine, KNN), (II) unsupervised learning (k-Means clustering) to achieve highest accuracy rate and best fit for the purpose.
* Designed a deep learning system which automates and abstracts away different repeatable routines using Python, TensorFlow, Numpy, Open CV.
* Developed an optical character recognition application for processing orders and invoices which minimalizes human error using Python, Tesseract OCR, GCP Vision API’s and Automation Anywhere.
* Improved In-house cognitive chatbot to understand and handle new business use cases by which ServiceNow requests were reduced by over 30% using Dialogflow and Python.
* Developed statistical model for predicting occurrence of next events using combination of ARIMA and Facebook’s Prophet with Python, Numpy, SciPy and Pandas.
* Performed Exploratory Data Analysis (EDA) for maximizing insights from the dataset also to detect outliers and extract important variables by graphical and numerical visualizations.
* Installed and configured Hadoop MapReduce and Developed multiple MapReduce jobs in python for data cleansing and pre-processing.
* Developed and designed solutions using with Spark in Python with Pyspark for SQL quires.
* Developed data extraction pipelines using AWS data pipelines to retrieve data from S3 buckets as well as, analyzing data with Amazon Redshift Spectrum to query terabytes of data present in S3.

**Environment: Python 3.6, Anaconda 4.6, Jupyter, NLTK, Keras, Google’s Tesseract OCR, Azure ML Studio, Azure CosmoDB, Google Cloud Platform DLP, Google Cloud Platform Vision API, IPsoft Amelia, Automation Anywhere**

**Client: AT&T, Bothell, WA November 2016 – November 2018**

**Role: Python Developer, Machine learning – Customer service**

**Description**: AT&T is an American multinational conglomerate holding company, where I worked as a part of RPA team. My role was to work with customer emails text data, detect patterns and trends from data and use those patterns to build a Python system with machine learning models to accurately separate signatures and extract important information. I also developed multiple classification methods for the company’s search engine to figure out the intent behind the search queries.

**Responsibilities**:

* Worked with Python libraries like NumPy, Pandas, Matplotlib, and urllib2 to extract data and manipulate data for exploratory analysis using data munging.
* Built models using Statistical techniques like Bayesian HMM and Machine Learning classification models like SVM, and Random Forest.
* Participated in all phases of data mining, data cleaning, data collection, developing models, validation, visualization and performed Gap analysis.
* A highly immersive Data science program involving data Manipulation & Visualization, Web Scraping, Machine Learning, Python programming, SQL, GIT, MongoDB, Hadoop.
* Worked with Spark Data frames, Spark SQL, Spark MLlib extensively for developing and designing POC's.
* Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, Naive Bayes, KNN.
* Perform extensive statistical computing and data manipulation using Hadoop Ecosystem.
* Applied Natural Language Processing (NLP) techniques to process the raw text.
* Validated the machine learning classifiers using ROC Curves and Lift Charts.
* Wrote Python routines to log into the websites and fetch data for selected options.
* Generated Python Django Forms to record data, login, and sign up of users.
* Developed data model, SQL Queries, SQL Query tuning process and Schemas.
* Involved in designing database Model, APIs, Views using python to build interactive web-based solution.

**Environment: Python, R, Django, MYSQL, Hadoop, Tableau, Linux, Machine Learning, NLP, Pandas, Numpy, Seaborn, Scikit-learn, TensorFlow, Keras, SAS, regression, logistic regression, random forest, SVM, NLTK, Classification.**

**Client: Creehan & Company, Canonsburg, PA November 2015 - July 2016**

**Role: Python Developer**

**Description:** Creehan & Company is an leading Healthcare technology company providing on-premises enterprise-wide software application platform and implementation services, where I was appointed to work with an application to capture clinical data formats, to calculate and submit clinical measures also assigned to work in the search module.

**Responsibilities**:

* Designed and developed the UI of the website using HTML, AJAX, CSS and JavaScript.
* Used Django Database API's to access database objects.
* Design, develop, test, deploy and maintain the web applications.
* Widely worked with R for statistical analysis and data visualisation using Ggplot2
* Wrote python scripts to parse XML documents and load the data in database.
* Rewrite existing Python/Django modules to deliver certain format of data.
* Used Selenium Library to write fully functioning test automation process that allowed the simulation of submitting different we request from multiple browser to web application.
* Used Python and Django to interface with the jQuery UI and manage the storage and deletion of content.
* Involved in development of Web Services using SOAP for sending and getting data from the external interface in the XML format.
* Responsible for debugging the project monitored on JIRA (Agile)
* Created unit test/regression test framework for working/new code
* Using Subversion control tool to coordinate team-development.
* Developed entire frontend and backend modules using Python on Django Web Framework.
* Responsible for debugging and troubleshooting the web application
* Created data access using SQL and PL/SQL stored procedures.
* Written many programs to parse excel file and process many user data with data validations

**Environment: Python 2.6, SciPy, Pandas, Spark, Bugzilla, SVN, C++, Java, jQuery, MS SQL, Visual Basic, Linux, Eclipse, Java Script, XML, JASPER, PL/SQL, Bootstrap, Oracle 9i, Shell Scripting, HTML5/CSS, Apache.**

**Client: Experian, Costa Mesa, CA May 2014 – October 2015**

**Role: Jr. Python/AWS developer**

**Description:** Experian is an American consumer credit reporting agency working with over a billion individuals and business, where I contributed in cleaning, formatting and pre-processing consumer data with maintaining the cloud architecture for optimizing the use of AWS services.

**Responsibilities**:

* Used Python with Django and Zope frameworks to develop applications.
* Implemented business logic using Python 2.7.
* Used HTML, CSS, JavaScript and jQuery and Implemented AJAX for dynamic functionality of a webpages for front end applications.
* Worked on front end frame works like Bootstrap for responsive webpages.
* Managed large datasets using Pandas data frames and MySQL.
* Developed and executed various MySQL database queries from python using Python-MySQL connector and MySQL database package.
* Implemented SQL Alchemy which is a python ORM library for complete access over SQL.
* Designed object model, data model, tables, constraints, necessary stored procedures, functions, triggers, and packages for Oracle Database.
* Fetched twitter feeds for certain important keywords using python-twitter library.
* Used Python library Beautiful Soup for web scrapping.
* Worked on Python OpenStack APIs and used NumPy for Numerical analysis.
* Used AJAX and jQuery for transmitting JSON data objects between frontend and controllers.
* Designed and managed API system deployment using fast http server and Amazon AWS architecture.
* Various static/media files along with database were hosted into Amazon cloud and support for Amazon AWS S3 and RDS was added.
* Implemented monitoring and established best practices around using elastic search.
* Install and configuring monitoring scripts for AWS EC2 instances.
* Worked with the design team for project structure.
* Created a Git repository and added the project to GitHub.
* Wrote Python scripts to parse XML and JSON documents and load the data in database.
* All projects adhered to agile principles to deliver commitments, meet acceptance criteria and ensure fast releases.
* Handled all the client-side validation using JavaScript.

**Environment: Linux, Windows, Python, Django, Zope, HTML, JavaScript, CSS, jQuery, AJAX and BOOTSTRAP, REST, MySQL, AWS, Elastic Search, SQL, Apache, Ubuntu.**

**Education:**

**Texas Wesleyan University, Fort Worth, Texas**

**Bachelor of Science in Computer Science**