# CAREER PROFILE

A dynamic and professional Data Engineer with excellent knowledge in analyzing large datasets, building, deploying, and Maintaining data support tools. Expertise in various technologies used for ETL Process & Data Analitycs with utilizing technologies like Python, Spacy, Apache Airflow. Possess excellent domain knowledge and leadership skills to handle the team as well.



# Dhamodharan K

Tech Lead | Data Analytics

# EXPERIENCES

Lead I - Software Engineering (Data Analytics)

Aug 2019 - Till date

UST Global, Bangalore (Client, Merck KGaA Bangalore)

As a member of the Data Analytics team, responsible for industrializing data feeds, creating pipelines into existing data systems, and facilitating connectivity of data and applied technologies between both external and internal data sources.

- Handling complete life cycle of Data Collection, Data Processing, Data Analysis for a given set of business problems
- Hands-on experience in designing and developing customized Frameworks for effective Data Extraction and Analysis
- Hands-on experience in various automation activities using Python & Browser Automation with using PySelenium
- Well experienced in converting the business problems into workable technical Prototypes
- Hands-on experience in implementing custom NER methods to extract the entities from biomedical text data
- Experience in developing and configuring the Apache Airflow based Pipelines
- Experience in developing API Services and Front-end Applications using Flask and Django
- Good understanding of deploying and managing the services on Linux machines
- Deployed services like Apache Airflow, label Studio, Django, and flask
- Experience in working with multiple databases like MySQL, Mongo, PostgreSQL
- Working experience in Agile & Ticket based environments like Jira, bitbucket and Service Now
- Experience in project tracking and team managements & Extensive knowledge of data sources like PubMed, Scopus, clinicaltrials.org, etc
- Alongside, Exploring the methods and technologies for improving and optimizing the solutions provided or for the methods implemented earlier

dhamodharanrk@gmail.com (mailto:dhamodharanrk@gmail.com)

**J** +91 98940 28181 (tel:+91 98940 28181)

www.dhamodharanrk.com (http://www.dhamodharanrk.com)

#### **EDUCATION**

**BE Computer Science and Engineering** Kumaraguru College of Technology, Coimbatore

2012 - 2015

**Diploma in Computer Engineering** PSG Polytechnic College, Coimbatore 2009 - 2012

# **TECHNICAL SKILLS**

Python, Shell Programming, HTML, jinja2

MySQL, MongoDB, PostgreSQL, Django, Flask, Spacy, Plotly, Pandas, Pyspark, beautifulsoup4, and More

## **LANGUAGES**

Tamil (Native)

English (Professional)

#### **INTERESTS**

Tutorial Writing (https://www.dhamodharanrk.com/portfolio/

Senior Software Developer (Python)

Blogs

Feb 2019 - Aug 2019

(https://sites.google.com/view/way2learning

### PPTS Pvt Ltd, Coimbatore

As a part of the Data Management and Python Automation Team, Supported multiple automation projects to solve business problems and Developed methods to automate end-to-end data extraction pipelines for effective Data Retrieval and Management. Also, responsible for handling the client requirements and managing the automation team

- Handled Data extraction and Analysis for Legal Domain
- Developed and deployed UI Based Data extraction Platform for effective data retrieval from various websites
- Implemented dynamic, template-based Python Codes to reduce development efforts
- Experience in OpenCV and Pillow libraries for Image processing as supported front-end Team for detecting and Croping the human faces from the profile picture
- Implemented Web Browser based Automation using Node JS, Selenium to perform various operations data mining automation
- Implemented API Services to provide data services to other Applications & teams
- Developed custom NER method to extraction user-related information like Name, Qualification, Job Title using Spacy
- With all together, improved the production speed of developing new domains from 4 per day to 10 per day
- Leading and Training the Team Members for the latest improvements in the technology front.
- · Handling multiple independent projects and Team Projects

# Software Engineer (Python)

Jan 2016 - Jan 2019

# Ugam Solutions Pvt Ltd, Coimbatore

As a part of the Infrastructure Development & Support team, Developing and managing data extraction for high complex websites and building futuristic platforms for effective data retrieval.

- Sound knowledge in e-commerce dataset
- Developed multiple data parsing methods using the custom framework
- Experiance in handling multilingual dataset & well experianced in most of the top e-commerce websites
- Worked on multiple automation tasks which include larger data processing, autometed data validation and QA
- Experince to handle the client deliveries by completing entire life cycle of the task like Scheduling,
   Monitoring, Exporting and Validating
- Implemented end-to-end automated solutions for projects having multiple stages
- Used multiple methods like NodeJs, Selenium for effective data retrieval and browser automation
- Closely worked with multiple technology teams for platform improvements and testing
- Done manual and automated code auditing and Data Auditing for Platform Migration from V1 to V2
- Provided continuous training for executive teams and the development teams for technology changes

### Domain Knowledge

Jan 2016 - Till Date

Summary of the domain related experiances

As a Python Programmer, solving and solved the challenging problems from the following areas

- E-commerce
  - Worked on custom developed Framework for Data Extraction based on Distributed Architecture
  - Developed more than 200 python Scripts for Data Extraction

- Experience in handling Multilingual datasets
- Experience in various end-to-end automation on top of the e-commerce dataset

#### Legal

- Sound knowledge in the legal dataset as worked on more than 300 websites
- Developed Dynamic & UI Based Data Mining Platform for effective data retrieval
- Handled automated data cleanings, validation, and matching

#### Bio-Informatics

- Sound knowledge in multiple data sources.
- Working experience in extracting data from various sources like PubMed, Clinical trials, Scopus,
   Grant, Patent, etc.
- Implemented methods to extract information like author, affiliation, and Products from text data using Python and SciSpacy
- Experience in handling custom NER models for extracting keywords using Spacy
- Understanding of author networks & Collaborations

#### News Sources

- Build a customized data mining platform with a stack of python libraries
- o Provides rich front-end for configuring the domains
- o Supports HTML, JSON, and RSS feeds
- Extracting and parsing data from around 500 Domains on a daily basis

## Official Projects

Jan 2016 - Till Date

Summary of the projects worked on

As Python Programmer, Worked on multiple individual projects and Team Projects

#### EDGE

- o Its an application to provide insights about the authors, institutes, and their research
- o Implemented end to end solution to extract and process the data from various data endpoints
- Technology Stack: Python, Pandas, MySQL, MongoDB, and other python libraries

# AIMS

- o It's a single source to monitor multiple BioMedical Data sources and get insights out of it.
- Designed and implemented pipelines to process data from various sources using a single
   Framework on top of Apache Airflow
- Developed a complete customized Framework to extract and process news data from various domains
- o Implemented common functionality used by all the team members.
- o Technology Stack: Python, Apache Airflow, Pandas, Spacy, and many more other libraries

# · Covid19 Dashboard

- Provides complete insights about the covid19
- Responsible for developing and providing support to data pipelines for the Dashboard
- o Technology Stack: Python, Pandas, Airflow, MySQL, MongoDB

# · Daily News Alerts

- o Extracting news from various sources and providing daily data delivery to multiple teams
- Responsible to automate end to end pipelines to ensure break free data delivery
- o Technology Stack: Python, Pandas, cron jobs, MongoDB

### Adhoc & Automation Projects

- Supported multiple Adhoc or one time projects from various teams
- Worked with multiple technologies such as Data Processing, Visualization, and Data Mining
- Technology Stack: Python, Pandas, cron jobs, MongoDB, MySQL, Matplotlib, Plotly, Sci Spacy,
   Spacy, and more

#### PPTS

Legal Data insights platform

- Supported to build an infrastructure to handle large datasets and extract data from plenty of sources
- o Technology Stack: Python, Pandas, cron jobs, MongoDB, MySQL, Spacy, Flask, and Django
- Ugam Solutions
  - Worked for extracting data from Amazon, Sears, Jet, Walmart, and so many e-commerce websites
  - Technology Stack: Python, Hbase, Hadoop, and other distributed concepts

# PERSONAL PROJECTS

Being curious about the latest technologies and passion towards programming,

MrSnippets (https://pypi.org/project/MrSnippets/) - A complete set of pre-defined code snippets to reduce development effort

MrCrop (https://pypi.org/project/MrCrop/) - Fast & accurate way of cropping faces from a batch of images in directory

Python Learning Platform (https://www.dhamodharanrk.com/lets\_learn/) - Simple inititive to provide easy learning solution for begineers

MrMap - Extract place names from a URL or text, and add context to those names

MrScraper - Framework to extract and process data with minimal coding experience

View More (https://www.dhamodharanrk.com/portfolio/projects/) - Website to showcase my working personal projects

# CERTIFICTAIONS

Diploma in Python Programing

Beginner to Advanced Python Programmer

Shaw Academy

Data Science using Python

Data Handling, Data Modelling & Machine Learning

Analytixlabs, Bangalore.

Diploma in Project Management - On Going Agile Project Management skills Shaw Academy

The Complete Python Pro Bootcamp for 2021 - On Going

To Learn the latest Tools and Technologies used for various Python Programming

Udemy

Spark and Python for Big Data with PySpark - On Going Spark Streaming, Machine Learning, Spark 2.0 DataFrames, and more Udemy



. ,	Program	

**Data Processing & Automation** 

ETL Pipelines (Custom Frameworks, Apache Airflow, Luigi)

Web Development(Flask, Django, Open API)

**Machine Learning Models & Libraries** 

Natural Language Processing (Spacy, NLP, Custom NER)

Database Technologies (MongoDB, SQL, PostgreSQL)