**PRADIP DHARAM**

Software Engineer - Python, Data • 14 years & 4 months [pradip.dharam@rediffmail.com](mailto:pradip.dharam@rediffmail.com) • Pune

[linkedin.com/in/pradipdharam](https://www.linkedin.com/in/pradipdharam/) • [github.com/pradipdharam](https://github.com/pradipdharam) +91-9975088871, 7972643894

**EXPERIENCE SUMMARY**

**Python** software development, **data & analytics** engineer. Proficient in writing **API contracts**, resulting in **python packages**.

**$50k plus savings** per annum basis python based static rules **anomaly detection** engine. Features development & refactoring of **AI ML** Document **Digitization** platform resulted in **reduced 80k plus manual touchpoints**. **Reduced** down the **efforts** by **40**% for the Business Operations Assurance, developed Order Fallout, Category & Recommendations **AI Prediction Engines**. **Aggregated** **100+ performance indicators** to data warehouses for **Scorecard**, General Ledger, **Consumer Lending**, Payment Integration, **Lease Management** etc. Developed **50+ data** **pipeline** **workflows** for Pharmaceutical giant to fulfill **business intelligence** needs.

**SKILLS**

**Coding**: Python 3, pandas, NumPy, Pydantic, pytest, git, Flask, Bitbucket, Github, Code Refactoring, PR Reviews, Products features API contracts development & refactoring, pyCharm IDE, SQL, Shell Scripting.

**Big Data, DB, ETL:** MongoDB, PostgreSQL, pySpark, ORACLE 10g, Teradata 12, DB2, Informatica Power Center 9.6.

**Machine Learning:** sklearn, NLP, SpaCy NER, Keras, Random Forest, k-means, xgBoost, MLFlow, 3-4 layered neural networks

**Advanced Beginner:** Airflow, Docker, Openshift, MLflow, Hive, Hadoop

**Novice To:** Kafka, CICD, Spark Streaming, Dagster

**Other:** Configured 0 code OpenTelemetry, COIN Spike on Provider Consumer to implement on Microservices

**WORK EXPERIENCE**

**Citicorp Services India Pvt. Ltd. 11 May 2020 – Present**

Assistant Vice President, Citi - Python Developer

* **GenAI EDGE Product**. Python 3, Pydantic, Code Refactoring, SQL Alchemy ORM, PostgreSQL, git, Microservices.
* Refactored python API contracts of EDGE Core based on jira tickets description. Dictionary related fast REST API’s to add related\_pipelines to response, SQL Alchemy ORM’s & Pydantic schemas to python packages.

Assistant Vice President, Citi - Python Developer, Data Engineer

* **AI/NLP Document Digitization** **Product**. Backend features development & refactoring. 20+ microservices containerised on Openshift ECS across 5 services. Monthly 20k+ documents processing, **reducing 80k+ manual touch points**, estimated 3 tenants viz. EMEA **Foreign Exchange** Non Deliverable Forwards, APAC Sanctions, Brazil Sanctions
* Refactored python API contracts to handle protected ZIP/PDF email ingestion, improved payload processing to split deal, & enhanced downstream XML generation per document. Regex driven entity extraction, from json generated by pyTesseract. Improved MongoDB performance with bulk inserts & updates. Designed JSON schemas to extract data from confirm PDF, single or multiple, capturing confirm number, description, signature (Y/N), and entity. Assign, approve, reject, payload side effect etc REST API’s refactoring.
* Tested Renovate PR changes across multiple product repos—service lib, connect lib, connect task, workflow service, and storage service. Integrated custom lib branches and image builds into services, verified on dev ECS. Merged 95% of PRs.
* Configured OpenTelemetry over Gunicorn for 5 services with 20+ Flask APIs, enabling traces, logs, and metrics from OpenShift containers to Grafana Observability. Used GitHub Copilot.
* Ensured 80%+ test coverage with pytest unit testing, and conducted PR reviews
* Assisted Product Architect, presented features to Product Owner and team, actively engaged in Agile ceremonies, and contributed to hiring by interviewing 5+ candidates.
* **Python** 3, **Code Refactoring**, **MongoDB**, **Git**, Flask REST Microservices, Bitbucket, **Airflow**, containers/pods monitoring on Openshift, PyCharm. UAT & PROD Support.

Officer, Citi - Python Developer, Data Engineer

* 13+ analysts stretching **900+ hrs** on month end closure **brought down to 200 hrs**. Rule Based **Anomaly Detection** Engine for data quality issues flagging, balance **reconciliations**, finding breaks, generation of custom excel reports**.** **Financial Risk Shared Services** EUC remediation for products viz. **Risk Weighted Assets**, **Derivatives**, After the Sale, **Over the Counter**, **Loans**, Debt, Cash & Due etc.
* **Python** processes developed to run business defined custom maintainable rules executed over the intermediate data generated after enriching the data on provided genesis source. Airflow tasks & dags. UAT & PROD support.
* Machine Learning (**NLP**) model prototype of **top 5 similar queries & resolutions recommendation**; using text **TF-IDF** vectorization & **cosine** **similarity**.
* **Python** 3, **Pandas**, **pySpark**, **Airflow**, Machine Learning Prototype.

**Tech Mahindra Ltd. 21 Sept 2016 – 01 May 2020**

Tech Lead - Python, Data Scientist

* **Order Fallout Prediction Engine** predicts likelihood of 25 tasks failures through which consumers placed order flows. Task features viz. geography, installation spec etc such 300 features were trimmed down to 56 to 96 based on feature importance. Insights presentations of confusion metrics for 25 task categories & bias. Prepared failure & reject reasons data set, **k-means** clustered & came up with failure categories based on business inputs. Finalized the prescriptive sequence of predicted task failures. Experimented **Random Forest**, **Naive Bayes, xgBoost** & 3 to 5 layered Neural Network**,** NLP**, K-means** clustering etc. For 10+ subdomains of **Business Operations Assurance**
* **Failures Category Prediction** & **Solution Recommendation Engine.** Text data such as prod support issue description, title, severity, time taken & solution logs commentary of L1, L2, L3 teams etc is sourced from AOTS tool. Defined custom categories based on k-means clustering. Top 5 solutions recommended for the predicted category.
* Raw data sets obtained from **Hive** tables on **HDFS** by **pySpark SQL** & from AOTS. Presented **Model & data insights** to Senior Scientists & Statisticians. Loaded **batch predictions** to HDFS or Oracle sink.
* Mainly **Stop words removal**, **lemmatization, POS** **Tagging**, **NER**,used to clean raw data; appropriate imputation used after analyzing the order task fields. **TF-IDF** vectorization for text, **One hot encoding** for categorical, **Standard Scalar** or **normalization** for numerical raw fields; used to engineer the vectors to train models.
* Broadband Operations metrics generated to warehouse based on business provided enrichments & transformations specifications on raw data, delta loads, and maintained providing production support.
* Tech stack is **Python**, **git**, **sklearn**, **keras**, **Pandas**, Numpy, **Data Wrangling**, Machine Learning, Supervised & unsupervised models training & experimentation. **Natural Language Processing**.

**YASH Technologies 15 Dec 2014 - 16 Sep 2016**

Sr. Software Engineer - Data Engineer ETL, Unix, SQL

* Data transformations workflow building for **Scorecard** domain for business metrics for regions viz. India, Nepal, America, Europe etc. CA Agile Central - Sprints. 100+ data pipelines productionized. 50+ KPI’s summarized to Scorecard data warehouse across levels viz. manager, group leader, branch, dealership, state, region, country, continent for business intelligence needs. Maintenance & production support.
* Informatica PowerCenter 9x, Unix Shell Scripting, DB2, SQL, **Agile**.

**Tata Consultancy Services 19 June 2013 - 12 Dec 2014**

Systems Engineer - Data Engineer ETL, Unix, SQL

* Bank’s data warehouse building for **General Ledger**, **Consumer Lending**, **Scorecard**, **Payment Integration**, **Lease Management** etc. for reporting requirements to ensure decision making, long term vision & business growth can be excelled. Data workflows, orchestration frameworks building. Unit & integration testing. Warranty & production support.
* Informatica PowerCenter 9x, **Unix Shell Scripting**, Oracle 10g, Greenplum etc.

**Cognizant Technology Solutions 24 Feb 2011 - 14 Jun 2013**

Programmer Analyst - Data Engineer ETL, Unix, SQL

* 50+ data pipelines workflows developed to build Divisional Data Warehouse for Pharma subdomains viz. **Hospital Demographics,** **Employees**, Samples, & Alignments etc. to fulfill reporting intelligence needs. Data engineering. Shell scripts were developed to drive. Extensively discussed with onsite counterparts, & came up with source to target mappings, unit tests. UAT & PROD support.
* Informatica PowerCenter 9x, **Unix Shell Scripting**, Oracle 10g SQL, **Teradata** 12, **SQL**

**TRAINING / COURSE**

**Low Level Design June 2024**

Design Principles (SOLID, KISS, DRY), Design Patterns

API Contracts Python Hands-on - ATM Design, Enquiry Handler, Scoreboard Notification etc.

**Big Data Masters Program certified by trendytech.in, Feb 2024**

Hands-on. Apache Hadoop, Hive, pySpark, Kafka, Airflow, Sqoop, Spark Streaming etc.

**Applied Machine Learning Course certified by** [**www.appliedaicourse.com**](http://www.appliedaicourse.com)**, Aug 2019**

Python Hands-on 15+ case studies, Machine Learning, NLP, Deep Learning, Random Forest, xgBoost, Regression, Naive Bayes etc.

**EDUCATION**

**Master of Science - Computer Science,** Grade: **First Class,** June 2008 - Jun 2010

New Arts, Science & Commerce College, Ahmednagar. University of Pune.

**Bachelor of Science - Computer Science,** Grade: **First Class**, June 2005 - May 2008

New Arts, Science & Commerce College, Ahmednagar. University of Pune.