Santosh Dighe
Solution Architect – Data & AI
E: sdighe1405@gmail.com ▪ M: +91 9004380181
Profile: https://santoshdighe.com ▪ Pune, India 411057
LinkedIn: https://www.linkedin.com/in/santoshdighe

SYNOPSIS

Over 15 years of experience in various phases of technology projects and business execution related to Data and Artificial Intelligence, including design, planning, product development, and business sales. Expertise in executing large-scale big data, machine learning, and data warehousing & analytics projects in varied industries, including retail, financial services, human resources, and insurance.

EXPERIENCE SUMMARY

- Strong problem solving, analytical, entrepreneurial, and communication skills with the ability to design and implement innovative technology solutions efficiently.
- Experience in business development, cloud, data transformation, computer vision, deep learning, and leading & managing technology effectively. Worked in major corporates in the U.S. and successfully used technology for business benefit.
- End-to-end Machine/Deep learning pipeline implementation with data collection, data preparation/labeling, model building, training, testing, and deployment with web hosting using technologies like Python, TensorFlow, Keras, PyTorch, OpenCV, and Data Augmentation.
- Expertise in implementing large-scale big data enterprise platform and analytics projects primarily using Cloudera CDH, Cloud, AWS, Hadoop Ecosystem, Apache Spark, Impala, Kafka, NoSQL MongoDB, and Linux technologies.
- Extensive expertise in implementing large-scale multi-dimensional Data Warehousing ETL projects primarily using PeopleSoft EPM, IBM InfoSphere DataStage, Unix, and relational databases in conjunction with various phases of the Software Development Life Cycle model.
- Proximity marketing using various location technologies like Bluetooth Low Energy (BLE) Beacons (Eddystone/iBeacon) and Geofencing, IoT, and location analytics. Web Development & hosting with AWS, REST API, Apache HTTP & Tomcat, Flask, VueJs, HTML, CSS, and Javascript.

CERTIFICATIONS/PROFESSIONAL DEVELOPMENT

Certification Program	Institute/Verified By	Month/Year
Neural Networks and Deep Learning	DeepLearning.AI/Coursera	Feb 2019
Big Data Using Hadoop	DePaul CDM IPD, Chicago, IL	Spring Quarter 2015
Data Science for Business	DePaul CDM IPD, Chicago, IL	Spring Quarter 2015
The Data Scientist's Toolbox	Johns Hopkins University/Coursera	Jan 2015
R Programming	Johns Hopkins University/Coursera	Feb 2015
Getting and Cleaning Data	Johns Hopkins University/Coursera	Mar 2015
IBM Big Data Fundamentals Professional v1	IBM/Prometric	Sep 2013

EDUCATION

Degree/Diploma	University / Institute	Description	Year
C-DAC Diploma	Institute for Advanced Computing and Software Development, Pune, India	PG Diploma in Advanced Computing (PG-DAC)	2006
B.E Computers	University of Pune, India	Bachelor of Computer Engineering	2004

TECHNICAL PROFICIENCIES

Type	Skills
Industries	Retail ▪ Finance & Banking ▪ Human Resource ▪ Insurance
Technologies	Deep Learning ▪ Python ▪ TensorFlow ▪ Keras ▪ PyTorch ▪ OpenCV ▪ Flask ▪ Cloudera CDH ▪ Amazon Web Services ▪ Hadoop Ecosystem ▪ Apache Spark ▪ Beacons ▪ IoT ▪ Hive/Impala ▪ Apache Kafka ▪ Apache Flume ▪ StreamSets ETL ▪ R ▪ Java ▪ HTML ▪ CSS ▪ Javascript ▪ Service Oriented Architecture ▪ IBM InfoSphere DataStage ▪ PeopleSoft EPM ▪ Data Warehousing ▪ ETL
Visualization/Modeling	Jupyter Notebook ▪ Matplotlib ▪ Plotly ▪ D3JS ▪ R ggplot ▪ Tableau
Version Controlling	GitHub ▪ Microsoft TFS ▪ PVCS ▪ SVN
Databases	MySQL ▪ NoSQL MongoDB ▪ PostgreSQL ▪ Oracle ▪ Microsoft SQL Server
Operating Systems	CentOS Linux ▪ Unix ▪ Mac ▪ Windows

WORK EXPERIENCE

PinCarts Private Limited   **Sep 2016 – Feb 2020**
Designation: Co-Founder & Technology Head – Data & AI

Description:
PinCarts is a technology-driven company with a vision to provide a seamless shopping experience to its valuable customers throughout their shopping journey using futuristic retail technologies like proximity marketing, personalization, advanced retail analytics, and indoor navigation. PinCarts' business model for brand retailers and shopping malls are meant to benefit the shoppers and retailers both.

Core Responsibilities:

- Strategic planning for the company's future growth, including business operations, marketing, and sales strategy development.
- Sales - On-boarded over 25 retail customers with end-to-end system integration for product inventories to provide digital and in-store navigation of products with personalized proximity marketing campaigns.
- Design and build efficient and secure centralized Enterprise Data Platform, mobile, and web cloud infrastructure using Amazon web services to support multiple workloads.
- Design, Develop and maintain advanced big data analytics and machine learning applications to support the company's vision of providing personalized and a real-time location-based shopping experience to our valued customers and operational efficiency for retailers.
- Proximity marketing strategy and implementation using various location technologies like Bluetooth Low Energy (BLE) Beacons (Eddystone/iBeacon) and Geofencing.

Solution Architect – Data & AI   Mar 2020 - Present

Core Responsibilities:
- End-to-end implementation & execution of Deep Learning projects with data collection, preparation/labeling, model building, training, evaluation, testing, and deployment.
- Setting up GPU Deep Learning infrastructure with NVIDIA CUDA Toolkit on Linux platform hosted on Amazon Web Services and Google Cloud Colab.
- Training and evaluating Deep Learning Convolutional Neural Network models for image object detection/recognition, classification, and segmentation problems.
- Model deployment with web hosting of trained models for real-time image object detection and classification using Python Flask web framework, REST API, HTML, CSS, and JavaScript.

Sample Projects Work (https://santoshdighe.com/#projects-section):

1. **Fashion Items Detection & Classification/Visual Recommendation System**
- Deep CNN Model in Python, Keras/TensorFlow for detection
- Web-deployed YOLOv3 based custom trained model with flask framework.
- Category level items detection & classification in real-time for product recommendations

2. **Social Distancing Monitoring System**
- Realtime tracking of crowded places with deep CNN for detecting people and violations
- PyTorch with Yolov5 and TensorFlow/Keras with Yolov3 & OpenCV implementations
- TensorFlow Object Detection Models (CenterNet/EfficientDet) implementation

3. **Human Pose Detection System**
- Deep CNN for person detection and Pose Guided Person Image Generation
- Used TensorFlow CenterNet HourGlass-512*512-kpts model.
- Multi-Person Pose Estimation & Keypoints detection

4. **Content-Based Fashion Recommendation System**
- Build historic data store in HDFS Hive/Impala from a relational database.
- Data preparation. Create similarity indexes using the Mahout RowSimilarity algorithm.
- Environment: Hadoop HDFS, Hive/Impala, Apache Spark, Mahout, Java, Python, & Elasticsearch.

Hexaware Technologies Inc.   **May 2006 – Jul 2016**
Designation: Technical Architect

Project: Global Accounting Hub and Operational Data Store
Client: Northern Trust Corporation, Chicago, IL, U.S.A
Work Location: Chicago, IL   Role: Technology Architect   May 2015 – Jul 2016

Environment:
Apache Hadoop, Apache Impala, Java, UNIX, Oracle Service Oriented Architecture (SOA) suite

- Implemented operational data store for accounting data in Enterprise Data Warehouse using Hadoop and Apache Impala/Hive metastore framework.
- Designed and Implemented service-oriented architecture for accounting business units.
- Integrated with upstream and downstream systems/applications with web services interface.

Project: Comprehensive Capital Analysis and Review (CCAR)
Client: Northern Trust Corporation, Chicago, IL, U.S.A
Work Location: Chicago, IL   Role: Automation Architect   Oct 2014 – Mar 2015

Environment:
IBM InfoSphere DataStage 8.5, IBM InfoSphere Server Manager and ISTOOL, WinSCP, UNIX, Oracle 11g

- Designed overall automation strategy for production execution of all CCAR schedules, which ensures fully automated execution and rerun ability and email notifications.
- Reduced time spent on quality assurance and decreased product defects by enhancing development procedures.

Project: Policy Administration System
Client: Trustmark Insurance, Lake Forest, IL, U. S. A
Work Location: Lake Forest, IL   Role: Data Architect   July 2014 – Sep 2014

Environment:
Microsoft SQL Server 2008, Visio Professional 2013

- Ensued database guidelines to create a new relational database in Microsoft SQL Server that will make all the data from the source data spreadsheets available to end-users.
- Standardized data storage and normalized the table/entity relationships up to a certain extent to eliminate data redundancy and manipulation anomalies.

Project: Global Data Warehouse (GDW)
Client: American International Group (AIG), U. S. A
Work Location: Mumbai/Pune, India   Role: Team Lead   Nov 2011 – March 2014

Environment:
IBM InfoSphere DataStage 8.5, PeopleSoft EPM 8.9, UNIX, Oracle 11g, PeopleSoft Scheduler

- Delivered a First-of-a-kind multi-dimensional data warehouse for HR analytics and operational reporting with IBM InfoSphere DataStage and Oracle as a back-end database in Unix environment.
- Automated and scheduled data load processes using Unix shell scripts and PeopleSoft scheduler with interactive email notifications and error log distributions.

Project: Northern Financial Objective Reporting, and Management Project (nForm)
Client: Northern Trust Corporation, Chicago, IL, U.S.A
Work Location: Chicago, IL   Role: Team Lead   July 2006 – Oct 2011

Environment:
PeopleSoft EPM 8.9, IBM InfoSphere DataStage, Windows XP 2002, UNIX, Oracle 10g, Control-m

- Implemented multi-dimensional financial data warehouse with IBM InfoSphere DataStage 7.5 and Oracle 11g as a back-end database in the Unix environment.
- Spearheaded a development team of over 15 members disseminated across onsite and offshore locations. Schedule execution automation with rerun ability.