Lalu Prasad Lenka

Data Scientist
Bangalore | +916360394977 | laluprasad.lenka95@gmail.com | LinkedIn | GitHub | Portfolio

Professional Summary

Data Scientist and seasoned developer with a strong knowledge of programming languages, data structures, algorithms and machine learning. I like finding creative, simple yet effective solutions to business problems using a combination of math, programming and business skills. I am super passionate about the reach of technology and when it comes to conquering new frontiers, I want to be there at the forefront. I love working with exceptionally talented people, always learning discovering new ways to do things better.

Skills

Analytical Skills	:	Machine Learning, Deep Learning, Natural Language Processing, Time Series Analysis, Optimization, Predictive/Prescriptive Modelling
Languages	:	Python, C++, Java
DB Scripting	:	MySQL, Oracle, MongoDB
Data Science Tools	:	RapidMiner, Knime
ML Libraries/Frameworks	:	Tensorflow, Keras, PyTorch, Scikit-Learn, Pandas
Big Data Tools/Skills	:	MapR, Hadoop, Spark, Kafka

Professional Strengths

- Data Science – Data Preparation, Predictive Modeling, Validation and Production rollout.
- Strong grasp on Statistical Analysis along with Machine Learning and Data Mining.
- Excellent problem-solving skills & programming skills.

Work Experience

Junior Data Scientist | Aptus Data Labs | July 2018 – Present

Working as a Data Scientist on data preprocessing, Machine Learning modeling, Advanced Analytics and Operationalization of Analytics.

- **Responsibilities**
  - **Technical**
    - Designing, building and validating data models.
    - Creating an **end-to-end** pipeline for the business solution.
    - Testing the models to deploy in production.
    - Developing the required operators/ connectors for any tool.
    - Delivering multiple complex projects on time, and to the satisfaction of the customer for various clients across Pharma, Oil Industry, Retail, Manufacturing, Academic Institutes, etc.
  - **Business**
    - Involved in presenting the result to senior management.
    - Making the client understand the insights and value our solution brings on the table.

- **Projects Undertaken**
  - **Project Title: Big Data Analytics Platform Benchmarking**
    - **Tools and Technologies: MapR, Apache Spark, Apache Mahout, Scala, Java**
    - **Client:** Leading Research Institute, India
    - **Project Description:**
      - We set up a four-node Hadoop cluster from scratch, starting from writing shell scripts to prepare individual nodes to install MapR services and later extended it to an eight node cluster. The benchmarking included the following:
      - K-means clustering algorithm results & speed in both Apache Spark & Apache Mahout for a 3 Million samples dataset.
      - **TestDFSIO** sequential read/write results for 2 TB, 5 TB & 10 TB.
      - Eventually, we built an IoT/Streaming Analytics platform where data gets read from sensors through **MQTT** server, published to **Kafka**, processed in **Spark**, persisted in **MapR FS** and **Influx DB**. Finally, reporting is done using **Grafana**.

  - **Project Title: Named Entity Recognition Engine**
    - **Tools and Technologies: Python, Keras, Kivy**
    - **Client:** Multinational Pharmaceutical Company
    - **Project Description:**
      - Built an **LSTM** based **Named Entity Recognition** model based on IUPAC dataset to detect chemical names and medicines in a given text.
      - The aim of the project was to fill the blank spaces in a given document with some meaningful text/chemical names by detecting nearby chemical names.
      - The complete pipeline was automated to meet the business requirement & deployed as Kivy desktop app.

  - **Project Title: Shipping Chain Optimization**
    - **Tools and Technologies: Python, MySQL**
    - **Client:** Fortune Global 500 Oil & Gas Company
    - **Project Description:**
      - The project included a comprehensive optimization of the company's ship routing & scheduling operation. The problem was a combination of capacitated vehicle routing & constraint satisfaction problem.
      - After careful scrutinization, a robust data model was built which helped us creating shipping lanes satisfying all constraints and one in-house **combinatorial optimization** algorithm was developed to find the optimized schedule of routing which reduced the cost of operation substantially.
      - Another **Genetic algorithm** based **Artificial Intelligence** was developed which creates cost optimized and constraint satisfied schedules.
      - We are currently trying to build a **Supply Chain Optimization** platform on top of it.

  - **Project Title: Document Classification Engine**
    - **Tools and Technologies: Python, MySQL, Keras, RapidMiner**
    - **Client:** Multinational Pharmaceutical Company
    - **Project Description:**

- Built a Document Classification Model to classify given documents on the basis of techniques used to prepare the medicine.
- Used **TF-IDF** score to create feature vectors and used **Random Forest** algorithm to build the model.
- The model helped the client to categorize documents in a fraction of time of what they used to do earlier.

Data Science Intern | Aptus Data Labs | Jan 2018 – July 2018

Worked as a Data Science Intern mostly on **Exploratory Data Analysis** & **Time Series Analysis**.

- **Projects Undertaken**
  - **Project Title: Time Series Analysis**
    - **Tools and Technologies: Python, Power BI**
    - **Client:** Fortune Global 500 Oil & Gas Company
    - **Project Description:**
      - Implemented proof of concept for supply chain optimization project by creating a time series model to forecast the load (oil, gas) requirement at different ports based on historical data.
      - Did extensive exploratory data analysis & data preparation to clean the data and used Time series forecasting algorithms to forecast the demand for next year.

  - **Project Title: Document Migration**
    - **Tools and Technologies: RapidMiner, Python, PowerShell, Adobe SDK, MS-Word Macros**
    - **Client:** Multinational Pharmaceutical Company
    - **Project Description:**
      - The project involved automation of **Document Migration & Management** for a multinational pharmaceutical company from its then immediate major acquisition.
      - The migration included splitting of the document into a number of method files, extraction of formulas, creation of template files, etc. without affecting the formatting.
      - Created tools for sanity check like document comparison tool to visually analyze the difference in two almost similar documents.
      - Successfully automated the whole process and reduced manual effort to a staggering 1-2% of initial.

Machine Learning Intern | Tata Consultancy Services | Jun 2017 – July 2017

Worked on a project "Image Attribute Extraction" which includes extraction of text from product images and populate specific attributes with extracted text.

- **Projects Undertaken**
  - **Project Title: Image Attribute Extraction**
    - **Tools and Technologies: Python, Keras, OpenCV**
    - **Project Description:**
      - It required working with deep neural networks like Convolutional and Recurrent Neural Networks.
      - Developed a combined architecture of **CNN & RNN** to build an Optical Character Recognition engine for text recognition.
      - Implemented the complete model using **Keras** API.

Academic Projects

- **AI for chrome dinosaur (Final Year Major Project)**
  - **Tools and Technologies: Python, JavaScript**
  - **Project Description:**
    - The aim of the project was to build a **Neuroevolution** based Artificial Intelligence bot that can play Chrome's dinosaur game.
    - A simple **3-layer Neural Network** was used to map the inputs i.e. distance from obstacle, speed and size of obstacle to output i.e. keystrokes(up/down).
    - **Genetic algorithm** was used to optimize the weights and biases of Neural Network generation by generation.

- **Gesture to speech conversion**
  - **Tools and Technologies: Python, JavaScript, Leap Motion Controller**
  - **Project Description:**
    - The aim of the project was to build an application that could help speech & hearing-impaired people to communicate.
    - We collected gesture(Indian hand signs) data using Leap Motion Controller device and tried to build a classification model on top of it.
    - We mapped gestures to alphabets using our model & used Google's text-to-speech engine to convert it to speech.
    - In the final stage of project, we were able to map series of gestures to complete words.

Education

2014–2018 | BTech in Computer Science and Engineering | College Of Engineering and Technology, Bhubaneswar **|** CGPA – 8.72

2011-2013 | Higher Secondary Education (Class XII) | D.A.V Public School, Talcher **|** 91%

Honors and Awards

Excellence Award

Oct 2018 | Aptus Data Labs

Excellence award for recognition of outstanding performance, significant contribution and dedicated service to drive customer engagement and technology learning in the field of Analytics.

Winner in IEEE colloquium Bhubaneswar Subsection

Nov 2016 | IEEE Bhubaneswar

Won first prize for paper presentation at IIT Bhubaneswar **IEEE colloquium**.

CET Merit Scholarship

College of Engineering and Technology

Got CET **merit scholarship** every year of BTech being a meritorious student.

Publication | Research

Gesture to Speech Using Leap Motion Controller | See Publication

Mar 5, 2017 | IEEE Delhi Section

This paper studies the possibilities of developing a gesture to speech and speech to text interface that uses Leap Motion sensor at its center for helping the substantial number of speech and hearing-impaired individuals (2.78% of total population) in our country.

Certifications

- Deep Learning Specialization from Coursera | See Certificate
- Machine Learning A-Z Hands-On Python in Data Science | See Certificate
- Data Science A-Z™: Real-Life Data Science | See Certificate
- Python: Design Patterns | See Certificate

I hereby declare that all the information above is true to the best of my knowledge and belief.

Lalu Prasad Lenka Bangalore, Karnataka