

Benazir de la Rosa - Data Scientist

E-mail: g1238gkl@gmail.com / Phone: 734-765-7190/ Michigan USA

Git: [nonameforpirate2 \(github.com\)](#) LinkedIn: [Benazir Abigail de la Rosa Muñoz | LinkedIn](#)

SUMMARY

- An Experienced Data Science and Data Analysis Application Systems Engineer/ Programmer in Design, Implementation, Machine Learning, Artificial Intelligence and Data Analytics Applications and Experienced with all phases of Software Development Life Cycles (SDLC).
- Translate real world problems into actionable advance analytics solutions. I support customers in terms of advance analytics applications in small and distributed scale (with spark/big data).
- Understanding data volumetry and how you can handle it depending on your application needs and size (pipeline development) and SQL, NoSQL technologies and advance analytics platforms, how they interact “under the hood”
- Manage the data in a way that I can choose the right database system or advance analytics platform according to user’s needs in case they need orientation about the data management.
- Experience developing data science application with Dataiku and python.
- Provide orientation about database systems in general, advance analytics solutions with an agile methodology, implement them according to the data volumetry and perform a good judge in terms of trade of budget vs optimal implementation.

Activities

Calculus all the flavors (UANL)	Artificial Intelligence (UANL / TU DELFT)	Recommender Systems (TU DELFT)
Controls Engineering (UANL)	Cloud Computing (TU DELFT)	Entrepreneurship Basic Course (TU DELFT)
Electronics Digital and Analogous (UANL)	Super Computing for Big Data (TU DELFT)	Modelling and Data Analysis in Complex Networks (TU DELFT)
Robotics (UANL-BTU)	Web Data Management (TU DELFT)	Software Architecture (TU DELFT)
Programming (UANL)	Machine Learning (STANDFORD)	Data Visualization (TU DELFT)
Data Bases and Data Mining (LEIDEN UNIVERSITEIT)	Cyber Data Analytics (TU DELFT)	Automation Engineering (BTU)
Spanish – Fluent	Nuclear Energy (BTU)	Electric Machines (UANL)
	English – Fluent	German - Fluent

SKILLS

Python, Spark, Scala, C#, R, JavaScript, Java Ubuntu/Linux Bash, SSH Entrepreneurship Courses Big Data Map and Reduce AWS Web Services (API's, Sagemaker, S3, Lambda, Glue, Athena) Snowflake * (Certification) Neo4j/Cypher	SAS * (Certification) Hadoop, Hive Troubleshooting Tableau, D3, Matplotlib Kafka, GraphX Time Series Analysis Numpy, Pandas, Scikit-Learn, XGBOOST, Scipy, NoSQL Architectures Kubeflow, Databricks	SQL Deep Learning GitHub Advance Excel Machine Learning GAccess C# Wonderware ETL pipelines Automation-Control Dev. Manufacturing – SCADA Technology – PLC's Arduino, RaspberryPi HTML/CSS
Dataiku * (Certification) Technology Integration	Agile/CRISP Jira/Confluence	

Education

- Autonomous University of Nuevo Leon Bachelor of Science in Mechatronics Engineering, Mexico 2009-2015
- Brandenburg University of Technology Cottbus (Exchange / Course Validation Bachelor's) Germany
- Delft University of Technology, Master in Computer Science & Data Science, The Netherlands

EXPERIENCE

Aug 2021 – Present American Century Investments Inc., Kansas City

Data Scientist / Data Engineer

Responsibilities:

- Worked with finance domain to develop new advance analytics solutions.
- My main duties consisted in use artificial intelligence (machine learning / deep learning) to solve business cases
- Use AWS technology to implement MLops life cycle. (Lambda, Athena, Sagemaker, Glue Crawlers)
- Use of agile version one (similar to Jira), document work with github and documents on project basis.

- Create model to suggest products to customers using graph database technology. Build neo4j network interaction and create cypher queries to do recommendations based on network contacts.
- Create model to classify phone calls base on type.
- Create model to predict Asset behavior based on NLP, macro economic information and weather.
- Implement use of Power Automate to automate reports for portfolio optimizers.
- Work with a fraud model.
- Use of tableau to do exploratory analysis over data and create dashboards for advance analytics solutions.

General Machine Learning	General Evaluation Cross Validation	GitHub
Anomaly Detection	Map and Reduce / Pyspark	Sagemaker, Lambda, S3, Redshift
Version One	Tableau	Athena
Python Code	SQL	Matplotlib Visualizations

Mar 2021 – Aug 2021 Anthem Inc, Chicago Illinois

Data Scientist / Data Engineer

Responsibilities:

- Worked with healthcare domain to predict medical circumstances
- My main duties consisted in use artificial intelligence (machine learning / deep learning) to solve business cases
- Use of Cloudera Hive/Hue platform to interact with customers data
- Use of Jupyter Notebooks on server with Pyspark to build Analytical models
- Build advance analytics Model. One model to predict future carrier usage on customer and another model to predict whether the claim should be paid by the company or should be paid by another third-party corporation. Support with large data pipelines.
- Extensive use of Hive/Pyspark/Python and Machine Learning, generate pickle file for production usage
- Use of AutoML platform to generate advance analytics models

- Support with python development code for common utility related to bias analysis on advance analytics models to be use for all data scientist across the company
- Use of bit bucket for code version control
- Use of Confluence to document model results
- Use of Jira for agile work methodology
- Use of Tableau to present information to Business area

General Machine Learning	General Evaluation Cross Validation	Bitbucket Control Version
Anomaly Detection	Map and Reduce / Pyspark	AutoML
Jira/Agile/Confluence	Tableau	Cloudera/Hue/Hive
Python Code	SQL / Hive / Oracle	Matplotlib Visualizations

Sep 2019 – Mar 2021 Raken Data Group, Miami Florida

Data Scientist / Data Engineer

Responsibilities:

- Worked with the data science team in several projects related to advance analytics solutions.
- My main duties consisted in use artificial intelligence (machine learning / deep learning) to solve business cases for different companies.
- Depending on the project, implement big data solutions in combination with data science to approach effective business cases solutions. For visualization purposes the use of tableau.
- In general, the skills used at the company are the next ones.
- Experience with Dataiku tool using different sources to extract the data and loading the data to various types of targets.
- Involved in the testing of Dataiku, an autonomous Machine Learning tool.
- Working with IT and data science teams, created an architecture and supporting processes for an edge node implementation of an advanced data science platform - Dataiku Data Science Studio (DSS)

- Discuss with companies advance analytics solutions by means of showing them as an opportunity offer PoCs where customers were able to foresee the importance of the correct use of artificial intelligence within their business case and later turn the negotiation in a short/long term advance analytics project. Using Dataiku.
- Optimized advance analytics workflows to save time and cost up to 100-300% (reduce by half or third part, the time needed to process de advance analytics pipelines). Using Dataiku.
- Write SQL queries to create aggregate information to feed machine learning models. (Hive, Oracle, Snowflake or Dataiku SQL recipes)
- Support with Snowflake development, setting up user permissions, create views, materialized views, udfs, store procedures and define warehouse size.
- Use of Tableau to present visual insights for customers.
- Provide training services for companies in Tableau tool (I gave two courses for DeAcero Company in Mexico and Colmena Company in Chile).
- Support to create pipelines invoking user define functions from snowflake.
- Use of pyspark and spark to aggregate data to feed machine learning projects.
- Use python code to create artificial intelligence models to be use in production environment. Ability to write production code using Dataiku.
- Use of optimization techniques for proof of concept for a bank. To select best outcome of asset combination by taking into account economic parameters over time.
- The information is based on projects developed at a bank, insurance company, steel retailer, light bulb company, coffee to go company and convenience store:

General Machine Learning	General Evaluation Cross Validation	Git Control Version
General Deep Learning	Map and Reduce	Dataiku
Fraud detection	Scheduling production code with Oozie	HQL
Anomaly detection	Write production code with spark	Tableau Visualizations (Include connections to other platforms such as Snowflake)
Time Series Analysis	SAS, Python, R	Amazon Web Services (Boto2, S3 bucket, Sagemaker, EMR etc...)

General Descriptive Analytics	SCRUM/Agile Methodology	Cloudera
NLP	Link business case according with statistic properties	Data Completeness analysis (Tell the customer whether the data is enough to work or he needs to input missing values)
Figure out general data behavior (trends, seasonality, median, mean, distribution type)	Selecting the correct data science measure according to the problem implementation	Recommender systems ArcGIS
Big Data	General data modelling	Snowflake Administration Configuration, Data Aggregation
Scripting, Structure and Unstructured data	Churn Detection	API's

Projects within Raken:

- Bank Profile Segmentation Mexico: Project to find internal fraud within the company. I gave support in the customer segmentation process, fraud detection and to help in the development of PySpark production code to run on a Cloudera cluster. Use of HQL to generate data aggregations. We schedule the task with Oozie. Project with CRISP methodology. Testing done with Dataiku. Use of R to get statistical properties from time series analysis.
- Insurance company Chile: Data analytics technology change. Migration from SAS to Dataiku. Use of snowflake to translate data analytics process with Dataiku. Main goal to estimate the price of a new insurance plan. Use of python to generate sampling methods and estimate prices for the new insurance plans. Project with Agile methodology. Application developed with Dataiku and Snowflake. Use of python to interact with some data aggregations and estimate new prices. Use of SQL in Snowflake to create data transformations, user define functions and stored procedures in Snowflake.
- Light bulb company, Stock Inventory Mexico: Use advance analytics models to predict product demand each month. Use python for machine learning purposes and data aggregation. Project with CRISP methodology. Application developed with Dataiku and Snowflake. Use of SQL code in Snowflake to create data transformations and python code to create the advance analytics pipeline. Use of R to obtain statistical attributes over time series analysis.

- Steel company, Recommender System Argentina/Mexico: Proof of concept made to the company to create a recommender system to suggest customers products that they would like to buy on their website based on previous purchases. Use of python to create several types of recommender systems and data aggregation purposes. Application developed with Dataiku and AWS. Use of python to create different recommender systems and retrieve data from S3 bucket aws.
- Convenience store, Reward Program Mexico (Orientation): Demonstrate the company leaders how they can create a reward program based on the product selling which might benefit the company the most by attracting new customers and getting the most out of the product waste. All of this with the correct implementation of artificial intelligence and data processing technologies. They are pretending to sell products based on points earned with purchases by the customer.
- PoC presented for a bank in Peru. People were looking forward to seeing what they can do with the recordings available on Phone calls. The proof-of-concept based on Natural Language Processing design was presented, consisted in transform the data from speech to text and then translate into sentiment analysis, NER (Named Entity Recognition) and some extra audio/call properties (tone, amount of novelty points, call duration), the goal to detect whether the call was bad quality or good quality, AWS technology was used in the proof of concept.

Jan 2016 – July 2017 Atos IT Solutions, Monterrey Mexico

Data Scientist / Software Developer

Responsibilities:

- Use data aggregation with excel.
- Use data aggregation with SQL.
- Control interaction between sql, excel and C#.
- C# programming for the purpose of api GRAccess interaction.
- Update information with a SCADA system Wonderware.
- Support to create front-end to update information automatically from wonderware 2012 to wonderware 2014.

Aug 2014 – Jan 2016 LEGO, Monterrey Mexico

Automation Engineer

Responsibilities: Technical support for manufacturing processes, scheduling installations for new production lines, development production improvements, AI in manufacturing (general production optimization).

PLC programming Allen Bradley	Production optimization on Manufacturing technology
Troubleshooting	C# Programming to monitor production in one area with old machines, SQL Connection.
Tech-Support	Use of Laser bars sick to measure boxes with artificial intelligence
HMI Programming Allen Bradley	General use of proximity sensors
Several communication protocols (RS-232, Ethernet, Profibus)	General use of actuator controls (Drivers for motors, valves, cylinders)

Dec 2013 – Aug 2014 Daltile, Monterrey Mexico

Automation Engineer

Responsibilities: Technical support for manufacturing processes, scheduling installations for new production lines, development production improvements, AI in manufacturing (general production optimization).

PLC programming OMRON	Use of Drives to control motors
Troubleshooting	HMI Programming OMRON
Tech-Support	Sick sensors
General controls to monitor sensors and actuators (pumps, motors, light sensors, proximity sensors)	Wonderware (Help to develop part of the application use in the company)
Several communication protocols (RS-232, RS-485)	Production optimization on Manufacturing technology

Mar 2013-July2013 Fest AG, Berlin Germany

Automation Engineer

Responsibilities: Know-How of the consultancy process for automation engineering projects. Test user interface software and correct errors (SCADA System in a Coffee Company). General understanding of automation engineering and network connections between engineering devices.

PLC programming	Remote server connections
Troubleshooting	Ethernet communication protocol
Automation-X programming	

Sep 2008-Sep2011 Sitel, Monterrey, Mexico

Customer Service Agent

Responsibilities: Dealing with customers, answer phone calls, schedule moving service, answer questions related to their billing service, create deals and offer to do retention over the service.

Mar 2006-Sep 2008 D' Confianza, Monterrey, Mexico

Jewelry Administration Sales

Responsibilities: Take care of product stock, do an overall register of sales, make sure the business infrastructure was in order.

University Projects Related to Data Science

- **Movie Recommender System.** We developed a movie recommender system with movie posters. We use OpenCV to decompose the images into their color properties (RGB), with the subtracted data and a distance metric build a similarity matrix. Afterwards, retrieve the image that is closer to the properties of the desire movie.
- **Rock Music Mixer System.** We download a dataset with peaces of classic rock songs. We processed the audio recordings with Librosa to extract audio properties. My mainly task was to process the dataset to identify novelty points in order to find out which was the best time to introduce the new song.
- **Text processing.** Use text to extract information out of it using map and reduce operations such as word count. Additionally, we gathered tweets and identify Hashtags to group tweets with the same topic. Profile Big Data System to understand the cost of hardware selection.
- **Actors Network Analysis.** We analyze how the friendships from a specific actor work, depending on the film where they interact with each other.
- **Bitcoin Network Analysis.** Download Bitcoin Data(10GB), out of the previous mentioned data a flowing money network was build. Thus, it was possible to infer properties from different accounts such as how many accounts are related to a given account, income money, outcome money, total balance in a period, historical total balance.

- **Artificial Intelligence Negotiation Agent.** Worked with an artificial intelligence framework to negotiate user preferences.
- **Nao Robot and Web Crawler Application.** We build an application with a NAO robot and a web crawler to read the newspapers to elderly people. The main ideas were the filter of the news by sentiment. (Good, Bad or Neutral)
- **New York Tourist Information.** Analyze New York city business are to identify daily activities that a tourist could enjoy. It was possible to see the distribution of business sectors around the city.
- **Medical Volume Visualization:** Reproduce with a Java Framework a 3D Object. It had the possibility to show different layers such as the surface or the skeleton.
- **Prototype to predict the Amsterdam Elections.** We gathered tweets from Amsterdam. Afterwards, we filter the information per political opinion (hashtags with political parties' names). My main task was to perform a correlation test over the geospatial data and with machine learning algorithm cluster the data.
- **Cancer Prediction in Gene.** Take Dataset with gene information build a network with Igraph (Python), extract properties for each gene and use the ground truth data to train a Machine Learning algorithm (SVM) to predict the cancer.

Others:

- ***Artificial Intelligence Conference May 2018(Berlin).** Pitch an idea about what can be done with the Sentinel-Hub data (Satellite Images). The proposed solution was **the agriculture resource manager**. The main goal of the solution was to identify a better distribution to produce fruits and vegetable per country and to create better routes for the transportation of food within the European Union.

Gaps:

Sep 2011 – Mar 2013 I was studying engineering in Mexico and Did exchange semester in Germany fulltime.

July 2013 – Dec 2013 I came back from Germany and I was studying fulltime engineering in Mexico.

July 2017 – Sep 2019 I was studying masters in Computer Science with Data Science specialization fulltime.