Work Portfolio

Wilmar German Fajardo Mendieta

February 7, 1992. Colombian
Bogotá, Colombia. Zip 111611
(+57) 3194490967
wilmargfm1021@gmail.com
Wilmar Fajardo
wgfajardom
Orcid Profile

Summary and conventions

During my career I have participated actively in more than twenty projects related to data and analytics. These projects are very diverse in terms of scope, technologies, and impact. They allowed me to interact with a wide variety of industries. Thanks to these experiences I have been able to developed some skills such as *analytical thinking*, *assertive communication* with teamwork and clients, *development and deploy* of machine learning models on production environments, and expertise on *cloud-computing*.

In the next pages, it is found a detailed list of the projects in which I have been involved. They are sorted in chronological order from newest to oldest, and also in groups according to the company I was working on. For each project the following pieces of information are presented. In the first row appears its name and its score. The score evaluates the degree of difficulty of the project, being five stars the highest difficulty. In the second row, it is mentioned the name of the client and its country. In the third row, some keywords related to the project are given. The first keyword corresponds to the industry of the client. Lastly, the scope of the project and my role on it are briefly described.

For a quick reference, the most challenging projects so far are listed below (top-5):

* Forecast of Tasks and CapEx (PHP)
* Cognitive Capture
* Real-Time Analytics for the Mobile World Congress 2018
* Design and Implementation of the Chatbot SECOBOT
* Concept and Applications of the Impulsivity in Solar Flares

IBM Chile

Aug 2022 – *Current*

Migration to GCP (★★★½☆)

Banco de Chile (BCH). Chile

Finance · GCP · On-Premise · Migration · Bitbucket · Scrum · Data Engineering

Support to the migration team. Facilitate them data and logs. Besides, share the understanding of the processes running on top of the on-premise arquitecture. Monitor services used from Google Cloud Platform (GCP).

Jun 2022 – *Current*

Rotative Consumption (★★★★½)

Banco de Chile (BCH). Chile

Finance · GCP · BigQuery · Oracle · SQL · Migration · Data Engineering

Recreate a dataset which is used as input for financial risk models. The dataset joins information from products, customers, debts, and so on. The sources of information are historical databases from Oracle. The recreation is made on BigQuery.

Mar 2022 – Aug 2022

Migration from SAS to Python (MigSAS) (★★★★½)

Banco de Chile (BCH). Chile

Finance · Hadoop · Zeppelin · Pyspark · Control-M · Atlassian · Scrum · Shell · Production Deploy

Migrate processes from SAS to an on-premise architecture based on Hadoop. Processes may be executed on demand or following a schedule. The former were developed via Zeppelin running on a Spark cluster. The latter were added to grids of Control-M, which were triggered by generic bash Shells. One grid was modified by myselft and its changes were deployed in the production environment.

IDATA

Nov 2021 – Feb 2022

Design and Implementation of the Chatbot SECOBOT (★★★★½)

Colombia Compra Eficiente (CCE). Colombia

Government · Azure · Bot Framework Composer · Bot Service · Adaptive Cards · UX

Creation of a chatbot to answer the Frequently Asked Questions (FAQ) received by the helpdesk of the *Agencia Nacional de Contratación Pública (ANCP)*, which belongs to CCE. The chatbot was designed on Bot Framework Composer following an *Interactive Text Response* (ITR) approach. The user experience is supported by adaptive cards of different kinds. It is a public tool that is found on this website.

Oct 2021 – Nov 2021

Automatic Estimation of Tasks Duration (EPP) (★★★½☆)

Ecopetrol. Colombia

Oil&Gas · Azure · Databricks · Pyspark · DevOps · Data Lake · SQL · Scrum

Calculation of tasks duration from projects of Ecopetrol based on the historical behavior of similar tasks. Three different scenarios are taken into account, namely, optimistic, regular, and pesimistic planning. I led the project from both technical and commercial perspectives.

Jun 2021 – Jul 2021

Improvements to the Forecast of Tasks and CapEx (PHP2) (★★★★★)

Ecopetrol. Colombia

Oil&Gas · Azure · Databricks · Azure Machine Learning · Pyspark · ML · DevOps · Data Lake · SQL · Scrum

Manteinance and improvements to the machine learning models of a previous project (PHP). Participation in the scope of the Product Backlog. Leadership from the technical point of view, especially the back-end.

Jun 2021

Proof of Concept of Azure Machine Learning (★★★½☆)

Sura. Colombia

Insurance · Azure Machine Learning · ML Classification · Python

Migration from an Azure Function that classifies emails on-demand to a model on Azure Machine Learning available in a public endpoint that fulfills the same purpose.

Apr 2021 – Nov 2021

Migration of the Certificate of Rates (★★★★½)

Asociación Chilena de Seguridad (ACHS). Chile

Social Security · Azure · Databricks · Pyspark · Data Pipelines · Data Lake · DevOps · Data Engineering

Reproduction of the Certificate of Rates (CR) as is stated in the *Decreto Supremo 67*. The CR determines the contribution rate that companies must pay in order to insure their workers. The contribution rate is computed from the accident indicators of each company. Input data is migrated using Azure Data Factory from SAP BW to an Azure Data Lake of various layers. Then, data is processed on Databricks where the indicators of the CR are calculated. This project has a strong component of data engineering and governance.

IDATA

Jan 2021 – Feb 2021

Preventive Maintenance of Ball Mills (★★★★★)

Antofagasta Minerals. Chile

Mining · Azure Machine Learning · ML Regression · IoT · Python

A ball mill is a machine used in mining industry to separate rock from copper. A regression model based on telemetry data from ball mills is developed in order to predict when such machines need maintenance. The target variable is the fill level of the ball mill.

Dec 2020 – Mar 2021

Analysis of Catastrophic and Non-Catastrophic Risks (★★★★★)

Sura. Chile

Insurance · Azure · Databricks · Python · Data Pipelines · Azure Synapse · Automation

Migration of a calculation process of catastrophics risks, i.e. those derived from natural disasters, and also non-catastrophic risks. This was done for all the insurance policies of the different portfolios offered by Seguros Sura in Ibero-America. A stochastic analysis is performed to know the *Aggregated Loss Distribution* (ALD). Additionally, a critical data source used for reporting was migrated via Azure Synapse.

May 2020 – Jan 2021

Forecast of Tasks and CapEx (PHP) (★★★★★)

Ecopetrol. Colombia

Oil&Gas · Azure · Databricks · Azure Machine Learning · Pyspark · R · ML Classification · ML Time Series · DevOps · Containers · Kubernetes · Data Lake · SQL · Scrum · Production Deploy

Forecasting the expenses of the projects of Ecopetrol based on their own historical expenses (CapEx model, ML Time Series). Prediction of task completion, taking as reference the project schedules, based on the historical completion of similiar tasks (Tasks model, ML Classification). The latter model was saved in a container on Azure Kubernetes and was deployed on a public endpoint thanks to Azure Machine Learning. In this way, the tasks model can be consumed on-demand. This was the first project of its kind for both Ecopetrol and iData. It generates weekly savings of around 44000 USD, which corresponds roughly to 3700 working hours.

Apr 2020 – Jun 2021

Workshops about Azure (★★★★★)

Microsoft. Colombia

Technology · Azure · IT Training · Teaching · Cloud Computing

Experience giving technical training on Azure to IT teams of different companies. In total, there were given around ten workshops on topics like Databricks, Azure Machine Learning, *Analytics in-a-Day*, among others.

Jan 2020 – Apr 2020

Cognitive Capture (★★★★★)

4-72. Colombia

Postal Service · Azure · Real-Time Processing · Computer Vision · OCR · Azure Functions · Azure Logic App · Data Lake · Cosmos DB · Flask · NLP

Real-Time processing of the incoming international mail received by 4-72. When a package arrives to the main office of 4-72, it is put on a timing belt. Just after, a picture of the destination label is taken automatically and it is loaded to Azure Blob Storage. Then, two Azure services, Computer Vision and Functions (serverless compute), extract all of the information related to the recipient of the package. This information is stored on a SQL database. Later, it is consulted by a Vanderlande scanner to direct the package to the adequate exit of the timing belt, i.e., the one corresponding to its destination. The entire processing for a single package occurs in less than 20 seconds and it is replicated continuously for thousands of them. Some packages cannot be not processed automatically, then they are processed by using a web interface to manually extract the recipient variables. The web interface was designed on Flask by our team.

Dec 2019 – Apr 2020

Chatbot EMMA (★★★★★)

Dirección de Impuestos y Aduanas Nacionales (DIAN). Colombia

Government · Azure · Python · Azure Functions · SQL · NLP

Chatbot that allows the interaction via voice and text. It was built to answer the Frequently Asked Questions (FAQ) about tax collection carried out by DIAN. It is integrated to Google Assistant. Its internal logic works with NLP to extract the entities of the question (main variable, time interval, and type of aggregation). Such entities form the on-demand query that is sent to a SQL database for answering the question.

Dec 2019 – Jan 2020

Credit Recommendation Model (★★★★★)

Caja de compensación familiar de Antioquía (Comfama). Colombia

Social Security · Azure · Databricks · Pyspark · Python · ML recomendación

Development of credit recommendation models for clients of Comfama. Each client is considered as a vector. Similarities between clients were computed using cosine distance and Pearson coefficient.

IDATA

Oct 2019 – Dec 2019

Detection of Anonymous Users (★★★★½)

Avianca. Colombia

Air Transport · Azure · Databricks · Pyspark · Python · SQL

Cross-exploration of different relational and non-relational databases of Avianca for identifying anonymous users. These users usually began a flight reservation but did not finish it. Four databases were considered, namely, purchasers, CMR, Amadeus, and LifeMiles.

MOCA

Jun 2019

MOCA Academy (★★½☆☆)

MOCA. Spain

Technology · Human Resources · Teaching

Design of the syllabus for a data science course for junior candidates. It will be the foundation for the further course *Skills to become a Machine Learning Data Scientist* offered by Universidad Nacional de Colombia in 2020.

Dec 2018 – Mar 2019

Proofs of Concept about Big Data Technologies (★★★½☆)

Mitto & Neinver. Spain

Finance · Retail · Geolocation · WiFi · Big Data · Python · Reporting · Insights

Mitto: understand how users relate with brands via an analysis of historical transactions of Mitto debit cards.

Neinver: analysis of the mobility of users on indoor spaces using WiFi signals. The technical limit of detection was given by the *Received Signal Strenght Indicator* (RSSI).

Sep 2018 – Aug 2019

Public Callouts for Funding of Digital Projects (★★★★½)

Callouts in Europe and North America

Technology · Government · Funding · Geolocation · Smart Cities · Management · Commercial

Synchronicity (European Union): provide a digital service based on Internet-of-Things (IoT) for some european cities, in order to improve the quality of life of the population and the growth of the local economy.

Small Business Research Initiative (United Kingdom): initiatives that allow citizens to be part of the continuous monitoring of road infrastructure in Durham and Blaenau Gwent counties.

Civic Accelerator - Road assessment (Canada): improve data acquisition, monitoring, and quality assessment of road infrastructure in Guelph. Prediction of future maintenances.

Lac La Biche (Canada): web or mobile reporting platform of road infrastructure with an interactive geolocation of issues, allowing the upload of the respective evidences such as images or videos.

Sep 2018 – Oct 2018

Lead a hiring process (★★½☆☆)

MOCA. Spain

Technology · Human Resources · Management · Communication

Design and implement a hiring process for a junior data scientist position. I was in charge of the callout, its spread, the technical test, and interviews.

Mar 2018 – Jun 2018

A/B Testing and Conversion Rates (★★★★½)

Inditex-Zara. Spain

Retail · Digital Marketing · Geolocation · Geofencing · Beacons · BI · SQL · Tableau

Exploration of the added value generated by a digital geomarketing campaign. The evaluation is made via A/B testing with symmetrical groups. Conversion rates were measured for each group at different stages of the purchasing process, namely, outside the mall, close to the store, and at the cash register (*funneling*).

Jan 2018 – Jul 2018

Real-Time Analytics for the Mobile World Congress 2018 (★★★★½)

GSMA & Fira. Spain

Logistics · Digital Marketing · Geolocation · Geofencing · Beacons · WiFi · BI · Real-Time · Tableau

Real-Time study of the flow of people inside Fira for any event of the MWC 2018. Onsite mobility was tracked via bluetooth signals of beacon devices or via GPS data given by the MWC app. This app was also used to collect online interactions, such as likes, views, interests, and so on. In this way, it was possible to have a complete and updated status of the MWC, both onsite and online.

MOCA

May 2016 – May 2017

Segmentation of users according to their mobility patterns (★★★★½)

MOCA. Spain

Technology · Geolocation · Urban Mobility · Physical Modeling · Graphs · Python · Zipf's Law

Algorithm that recreates the main mobility patterns of users from their historical geolocation data. Then, users are segmentated according to similarities between thier mobility patterns. Such paths or patterns follow an exponential trend known in the literature as Zipf's law.

Propios y externos

Jun 2021 – Sep 2021

Lethal Wilt Prediction (★★★★½)

Fedepalma & Cenipalma (DS4A Capstone). Colombia

Agriculture · ML Time Series · Geospatial Analysis · Python · AWS · SQL · Dash

Lethal Wilt (LW) is a disease that may destroy large fields of oil palm crops. The eastern region of Colombia is particularly affected by LW. It is developed a machine learning model that predicts the appearance of the disease taking into account historical data of meteorological variables from the last n-months, being n a free parameter of the model. The prediction will allow to identify the outbreaks of infection spatially.

Feb 2019 – May 2021

Geofencing algorithm (★★★★½)

Own Project. Colombia

Technology · Geolocation · Geofencing · Urban Mobility · Python · Applied Physics · Geometry

Method to track when a user enters or exits a given geofence. It is calculated continuously how often (time interval) the geolocation of the user (mobile phone location) needs to be queried. The algotirhm depends exclusively on how fast the user moves (kinematics). The geofencing algorithm works for geofences of different geometries and sizes. This development can be applied to platforms of digital geomarketing.

Jun 2014 – Sep 2014

Theoretical and numerical spectra from blazars (★★★★½)

Own Project. USA & Colombia

Research · Astrophysics · High Energies · AGNs · Python

Theoretical and numerical description of blazars spectra. Blazars are *Active Galactic Nuclei* (AGNs) whose powerful jets point directly to the Earth. They are one the main sources of gamma rays in the universe. Two components of the spectrum were studied and modeled, low-energy synchrotron contribution, and high-energy inverse Compton effect emission.

Jun 2012 – Nov 2017

Concept and Applications of the Impulsivity in Solar Flares (★★★★★)

Own Project. Colombia

Research · Heliophysics · Satellital Data · Automatic Pipelines · Solar Flares · IDL-SSW · HXR · Microwaves

Research work beginning from the final years of my B.Sc. in physics and finishing in the last year of my M.Sc. in astronomy. In the first part of the study, it was defined the concept of impulsivity for solar flares and a method to measure it from HXR emissions (*impulsivity parameter*). More than one hundred solar flares were classified according to the such parameter. In the second part of the research, it was explored the joint effect of impulsivity and magnetic trapping, in order to determine the kinematics of electrons inside coronal loops just after a solar flare occurs. For this, HXR and microwaves data were used coming from satellital and ground-based observatories, respectively. The timerange of the worksample spans between the maxima of Solar Cycles 23 and 24 (2002-2013). An automatic pipeline was developed to process both kinds of data.