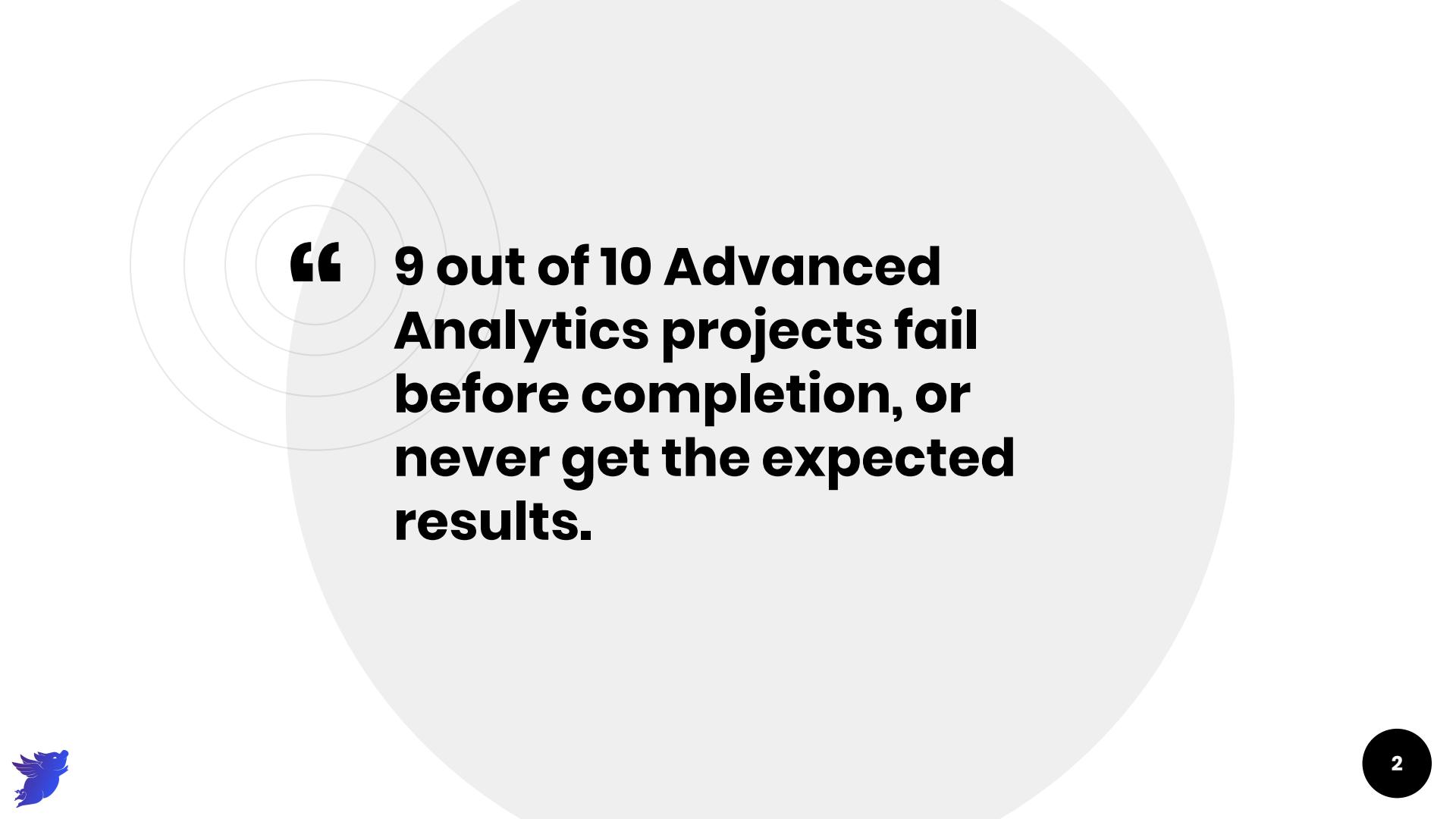
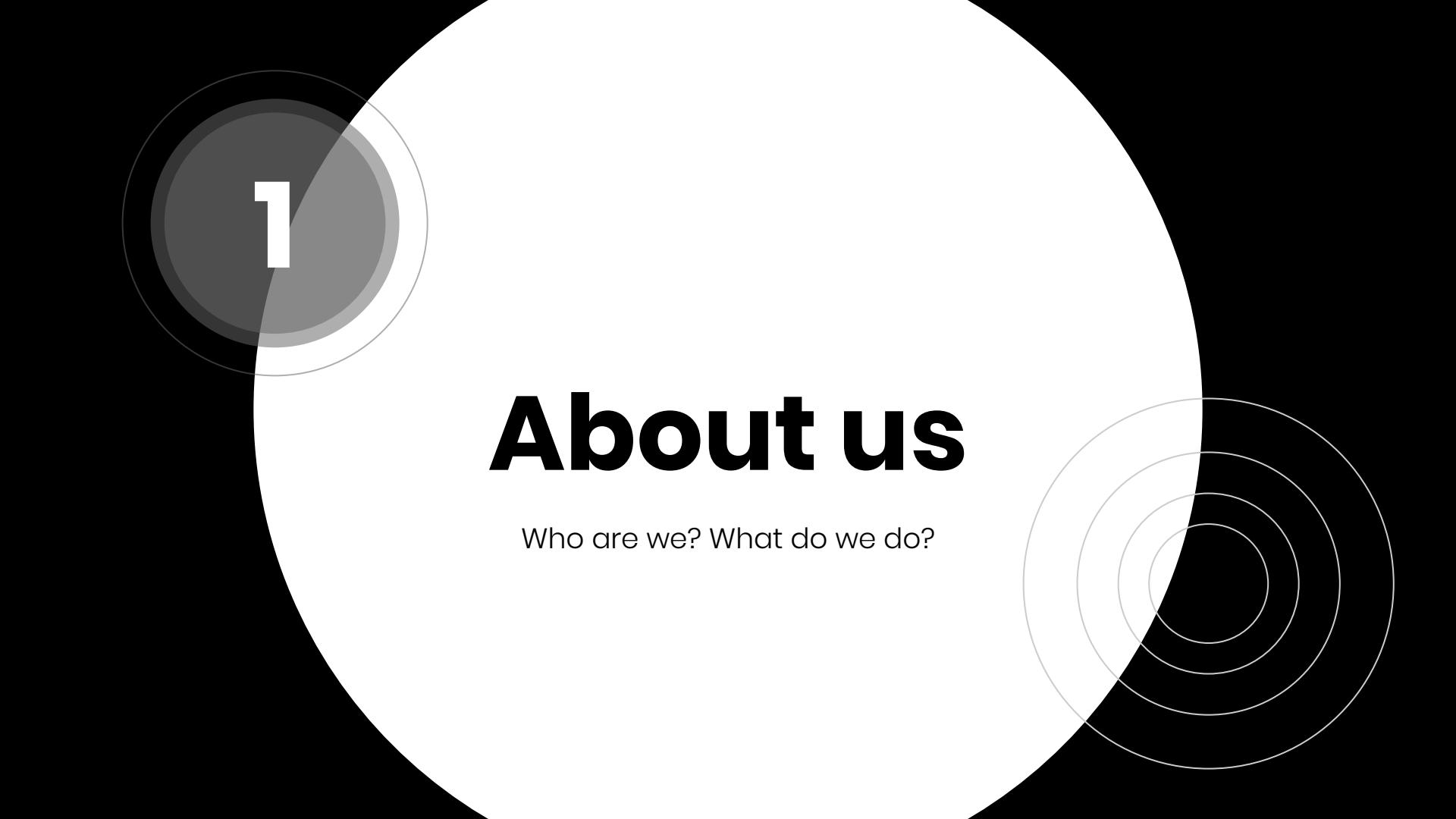


Advanced Analytics Portfolio





“ 9 out of 10 Advanced Analytics projects fail before completion, or never get the expected results.



1

About us

Who are we? What do we do?

Team

Pedro Muñoz

- Data Scientist
- Big Data Architect
- Aerospace Engineering & Master in Applied Maths
- Previously at Pragsis Bidoop, P3 Group

David A. Cañones

- Data Scientist
- 5 years of experience crafting models
- MS Industrial Engineering, MBA
- Previously at Pragsis Bidoop, Kernel Analytics (BCG Gamma)



2

Portfolio

Our Projects and Results



Energy

Power generation forecasting

Description

Prediction of future power generation of renewable energy plants based on meteorological information (three-dimensional grids).

Results

Improvement of 10 percentage points in the normalized forecasting error versus previous models. Millions of \$ in savings for our client when going to the US electricity bid.

Technology

Data ingestion with Spark (Scala) over Hortonworks infrastructure. Modeling with scikit-learn and lightGBM. Orchestration with Apache Airflow.



Predictive maintenance

Description

Calculation of damage due to fatigue of structural components of wind turbines and the time of failure forecasting.

Results

Development of a model which emulates the results of commercial enterprise software, which generates a saving of ~30K€ / year in software licenses and a huge increase in Know-How about wind turbines fatigue.

Technology

Treatment and study of fatigue time series with pandas.
Predictive model based on linear methods. High performance computing with NumPy and Fortran 90.





Telecommunications

Mobile network user experience

Description

Development of a Customer Experience Management (CEM) Framework. Network quality data ingestion from 3G / 4G antennas. Creation of a model to predict customer experience effect on churn and complaints.

Results

Complete mobile network monitoring and ability to identify antennas with low performance and negative impact on user experience at individual aggregation level.

Technology

Data ingestion made with Python and Impala over Cloudera infrastructure. Modeling with scikit-learn. Visualizations with Matplotlib and seaborn. Dashboards with Microsoft Power BI.



User complaints forecasting

Description

Predictive model for customers probability to open issues due to disagreements in billing, based on their consumption patterns and individual personal profile.

Results

Classification model with ~ 0.85 AUC over very unbalanced data.
Automation of part of the customer service process.

Technology

Creation of data mart using PySpark. Modeling using Spark MLlib.
Visualizations with Matplotlib, Plotly and Seaborn.



Mobile Operators Benchmarking

Description

Massive data ingestion of antenna performance, package loss, network parameters... collected by cars to calculate those KPIs associated to the benchmarking.

Results

KPIs of network performance for the mobile operators benchmarking on each country.

Technology

Spark on Java for the ETL, Impala and Hive for data analysis and Tableau for data visualization.





Healthcare & Pharma

Data pipelines optimization

Description

Client had a big problem in terms of Spark process performance due to the huge amount of data to process (~ TB).

Results

Complete optimization of the data pipeline that allowed to go from computing times in the order of days to a few minutes.

Technology

Spark with Scala for data processing. Flume and Sqoop for data ingestion. Storage in HDFS available through Hive SQL engine. All over MapR technology.



Analytical engine for researchers

Description

Project with support of the H2020 program. Creation of an analytical engine capable of scraping data from different sources and relating them to each other. Search engine with the use of Natural Language Processing.

Results

Application that is able to translate searches in Natural Language into exportable tables and visualizations made from heterogeneous data.

Technology

NLP model made with NLTK. Web scraping with Selenium, Beautiful Soup and Requests. Web application based on Django. Visualizations with Bokeh.



A photograph showing a row of shopping carts lined up in a supermarket aisle. The carts are made of metal wire and have red plastic wheels. The word "Retail" is overlaid in large white letters across the center of the image.

Retail

Footfall Analytics

Description

People detection and counting both in store and on public roads and heat maps to determine the areas with the greatest influx in supermarket shelves.

Results

Success in getting the best possible store location (pedestrian analysis) and the best product layout (store heatmaps).

Technology

OpenCV for image processing and TensorFlow for detection models. All restricted to the requirements of the Google Edge TPU hardware.



Classification of customer issues

Description

Classification of customer issues in online stores. Data processing using NLP and categorization into 18 different types.

Results

Automation of part of the customer support process. Only those issues that the algorithm is unable to classify with confidence are handled manually.

Technology

NLP model made with NLTK. Integration with the rest of the systems in the form of REST API using Flask.



A photograph of a person wearing a white skull mask and a red hoodie, sitting at a desk in a dimly lit office. They are looking at a laptop screen displaying code. The desk is cluttered with multiple monitors, keyboards, and electronic components. In the background, other people are working at their desks, and the office has large windows. The word "Startups" is overlaid in large white letters across the center of the image.

Startups

Moderator Guru

Description

The Moderator Guru is a Machine Learning, Natural Language Processing based text moderator service that detects and classifies offensive text messages.

Results

API endpoint capable to process thousands of requests per minute. Currently working on a Wordpress plugin.

URL: <https://moderator-guru.com>

Technology

NLP model focused on speed. Django for backend and vanilla frontend.



Renfe Guru

Description

Web scraping of renfe.com to extract high speed train data, in order to take advantage of pricing changes and save money in train journeys.

Results

Work In Progress. Expected alarm system for ticket pricing changes and Machine Learning model for predicting when these changes are more likely to happen.

Technology

Working on the most suitable algorithm for the train data. Django for backend and vanilla frontend.



DataTau

Description

DataTau is the reference newsboard for Data Scientists and Data Engineers inspired by the popular Hacker News. The site went down for a month and we decided to recover and open source it, rewriting application code from scratch and providing hosting.

Results

It reached Hacker News frontpage on its launch day and it is currently serving the hottest Data Science news to the World.

URL: <https://datatau.net>

Technology

Django for backend and vanilla frontend with a lot of Jinja..



