Name: Antoine Bertin
Age: 35 years old
Occupation: Data Scientist, Former Sommelier & Blockchain Enthusiast
Link to schedule a call: [https://calendly.com/antoinebertin/30](https://calendly.com/antoinebertin/30)
Email: [antoinebe35gmail.com](mailto:antoinebe35gmail.com)

Languages:

* French
* Spanish
* English

About me:
Hey there! I'm Antoine, a passionate data scientist with a business mindset. I dream of a world where solving challenges is as simple as running a pip install command.
Run '*pip install antoine*' now in your terminal and execute the command '*hire_antoine*'.

If you want to contribute to the 'antoine' package and make me better, check out my GitHub: [http://github.com/monolok/antoine](http://github.com/monolok/antoine)


Skills:
* Project Management
* Business Development
* Process Improvement
* Process Automation
* Cluster Analysis
* System Deployment
* Prediction
* Deep Learning
* Machine Learning
* Big Data
* Big Data Analytics
* Data Mining
* Data Modeling
* Algorithms
* Machine Learning Algorithms
* Artificial Intelligence (AI)
* Blockchain
* Non-Fungible Tokens (NFTs)
* Python (Programming Language)
* Web Development
* Databases


Experience:

AI Consultant & Product Owner at CloprNFT
March 2022 - March 2024 (2 years 1 month)
Worldwide

Clopr's web3 mining dApp enables one-click generation of token-bound derivatives using CloprBottle NFTs, adding utility to your NFTs. (<https://app.cloprnft.com/>)

My achievements:

* Coordinated a global tech pipeline across the USA, Georgia, France, and Argentina, leading teams to integrate blockchain and AI. Managed the lifecycle of the blockchain application from inception to launch, ensuring product cohesion.

* Devised a data collection strategy by working with Hollywood scriptwriters to develop a standardized story template for efficient data mining. This enabled effective AI model fine-tuning.
* Led the integration of AI technology with blockchain and spearheaded the deployment of smart contracts co-authored by myself. This resulted in the creation of a groundbreaking platform for generating token-bound derivatives within the NFT ecosystem.
* Crafted comprehensive whitepapers and technical specifications, streamlining regulatory compliance and legal clarity for users. This facilitated trademark registration and defined platform terms and conditions.
* Implemented streamlined feedback processes by automating user feedback mechanisms to encompass the entirety of the tech pipeline and beyond. This ensured that no valuable information was overlooked, providing insights crucial for understanding and prioritizing tech initiatives effectively.


Founder and Former CEO (Restaurant sold) at Galette Café Saint Germain
Self-employed · Sep 2012 - Dec 2023 · 11 yrs 4 mos · On-site
2 rue de l'université, Paris, France

"Galette Café" is nestled in the heart of Paris, embodying the essence of Breton cuisine. Our mission is to honor the culinary heritage of Brittany by showcasing the finest organic, artisanal products sourced from local producers.

My achievements:

* Built a Concept from Scratch: I conceptualized and established Galette Café, laying its foundation from the ground up.
* Built a Reputation: My dedication to excellence garnered us multiple prestigious awards, including the coveted "Best Restaurant" recognition from Trip Advisor.
* Built an Anti-Fragile Business: Amidst social and economic challenges, Galette Café stood resilient, navigating through uncertainties with grace.
* Not Afraid of Pivoting: Adapting to evolving consumer trends and technological advancements, I ensured Galette Café remained at the forefront of the market.
* Sustained Business Growth: Over a span of 12 years, I sustained consistent growth and profitability.
* Staff Cohesion: Fostering a cohesive team environment, I nurtured a dedicated workforce, achieving remarkable staff retention rates and fostering a positive workplace culture.


Old Managerial Experiences:

Food & Beverage at The Little Nell
Nov 2015 - Nov 2017 · 2 yrs 1 mo · Aspen, Colorado
Food & Beverage at InterContinental Hotels Group
Jan 2010 - Jan 2011 · 1 yr 1 mo · Nicaragua, Managua
Food & Beverage at The Little Nell
Nov 2008 - Nov 2009 · 1 yr 1 mo · Aspen, Colorado


Education:
Jedha Bootcamp
Master's degree, Artificial Intelligence
Grade: Full stack data science & engineering program grade: Full stack data science & engineering program
📌 **Data** & **Big Data**:

Databases: PostgreSQL, MySQL
Data transfer & cloud services: ETL | AWS | GCP
Web Scraping: BeautifulSoup, Scrapy
Big data: Spark, Databricks

📌 **EDA**:

Statistics and Probability
Data viz & Data mining
Libraries: Matplotlib, Seaborn, Plotly

📌 **ML** & **Deep learning model creation**:

Libraries: Pandas, Numpy, Scikit-Learn, Keras, TensorFlow, Spacy, Word2Vec, Langchain, HuggingFace
Supervised ML: Linear & Logistic Regressions, Ensemble learning, SVM
Unsupervised ML: Clustering algorithms with KMeans, DBscan
Dimensionality Reduction: PCA, LSA
Regularization: lasso, ridge, elastic net
Deep Learning: RNN, CNN, GRU & LSTM, GAN, transformers

📌 **Leveraging open source models**:

Transfer learning & Fine-tuning
Training big models with LoRA and QLoRA
RAG tasks: Vector databases with Pinecone, chroma

📌 **Deploy API** & **App**, **monitoring**:

Docker, MLflow, Amazon S3 & EC2, Streamlit, Heroku, FastAPI

📌 **Data** & **Big Data**: Databases: PostgreSQL, MySQL Data transfer & cloud services: ETL | AWS | GCP Web Scraping: BeautifulSoup, Scrapy Big data: Spark, Databricks

Skills: Machine Learning · Cluster Analysis · Big Data · Big Data Analytics · Artificial Intelligence (AI) · Deep Learning · Data · Algorithms · Data Modeling · Data Mining

Glion Institute of Higher Education
Bachelor of Applied Science (B.A.Sc.), Business Administration and Management · (2008 - 2012)
Universidad Europea
Grado en Administración y Dirección de Empresas, Business/Managerial Economics
Le wagon
Full stack - Ruby on rails, Computer Science
Les Roches Bluche International School of Hotel Management Bachelor of Applied Science - BASc

A fun fact is my high level of wine knowledge Certification:
Wine & Spirit Education TrustWine & Spirit Education Trust - WSET 3
Court of Master Sommeliers - CMS 2Court of Master Sommeliers - CMS 2

Projects and works - Machine and deep learning projects I have worked on, all the code can be found in my Github (*https://github.com/monolok/antoine*):

Big Data (Databricks and PySpark) - Steam's videogames platform Big Data (Databricks and PySpark) - Steam's videogames platform

Associated with Jedha BootcampAssociated with Jedha Bootcamp
I conducted a comprehensive analysis of the games available on Steam's marketplace as part of a project for Ubisoft. The goal was to gain insights into the video game ecosystem and current trends, ultimately aiming to understand the factors influencing a game's popularity or sales globally.

At a macro level, I delved into various aspects such as the distribution of games by publishers, the ratings of games, trends in release years including the impact of events like Covid-19, pricing strategies, language representation, and age restrictions.

For genre analysis, I explored the most prevalent genres, their review ratios, publishers' genre preferences, and the most lucrative genres.

Additionally, I examined platform preferences, determining whether games were predominantly available on Windows/Mac/Linux and if certain genres were more likely to be found on specific platforms.

To accomplish this, I utilized Databricks and PySpark for data manipulation and analysis. Leveraging PySpark's functionalities such as getField() and explode(), I handled the semi-structured dataset efficiently. Furthermore, I employed aggregate functions and groupBy to conduct segmented analysis, ensuring a comprehensive understanding of the data.

My deliverable includes one or several notebooks showcasing data manipulation with PySpark and data visualization using Databrick's dashboarding tool. I made use of the "publish" button on Databricks notebooks to create public URLs for easy access by the jury, adhering to the project guidelines. Links to the published notebooks are available in my GitHub repository for the jury's convenience.I conducted a comprehensive analysis of the games available on Steam's marketplace as part of a project for Ubisoft. The goal was to gain insights into the video game ecosystem and current trends, ultimately aiming to understand the factors influencing a game's popularity or sales globally. At a macro level, I delved into various aspects such as the distribution of games by publishers, the ratings of games, trends in release years including the impact of events like Covid-19, pricing strategies, language representation, and age restrictions. For genre analysis, I explored the most prevalent genres, their review ratios, publishers' genre preferences, and the most lucrative genres. Additionally, I examined platform preferences, determining whether games were predominantly available on Windows/Mac/Linux and if certain genres were more likely to be found on specific platforms. To accomplish this, I utilized Databricks and PySpark for data manipulation and analysis. Leveraging PySpark's functionalities such as getField() and explode(), I handled the semi-structured dataset efficiently. Furthermore, I employed aggregate functions and groupBy to conduct segmented analysis, ensuring a comprehensive understanding of the data. My deliverable includes one or several notebooks showcasing data manipulation with PySpark and data visualization using Databrick's dashboarding tool. I made use of the "publish" button on Databricks notebooks to create public URLs for easy access by the jury, adhering to the project guidelines. Links to the published notebooks are available in my GitHub repository for the jury's convenience.
Skills: Big Data · Machine LearningSkills: Big Data · Machine Learning
Deep Learning (Embedding) - Spam detector for AT&TDeep Learning (Embedding) - Spam detector for AT&T

Associated with Jedha BootcampAssociated with Jedha Bootcamp
I worked on a project with AT&T to address a major concern for their users: spam messages. Despite their manual efforts, AT&T sought an automated solution to flag spam and protect their users' experience.

In this project, I aimed to develop a spam detector solely based on SMS content. To begin, I utilized a provided dataset to train deep-learning models for this purpose.

Following the project scope, I adhered to some helpful guidelines:

1. Starting Simple: I focused on developing effective models without unnecessary complexity.

2. Transfer Learning: Given the limited dataset, I leveraged transfer learning to tap into the knowledge of pre-trained models trained on vast amounts of data.

The project's deliverables can be found in my GitHub and included:

- A notebook containing preprocessing steps and the training of one or more deep learning models.
- Clear documentation of the model's performance.I worked on a project with AT&T to address a major concern for their users: spam messages. Despite their manual efforts, AT&T sought an automated solution to flag spam and protect their users' experience. In this project, I aimed to develop a spam detector solely based on SMS content. To begin, I utilized a provided dataset to train deep-learning models for this purpose. Following the project scope, I adhered to some helpful guidelines: 1. Starting Simple: I focused on developing effective models without unnecessary complexity. 2. Transfer Learning: Given the limited dataset, I leveraged transfer learning to tap into the knowledge of pre-trained models trained on vast amounts of data. The project's deliverables can be found in my GitHub and included: - A notebook containing preprocessing steps and the training of one or more deep learning models. - Clear documentation of the model's performance.
Skills: Machine Learning · Deep Learning · Artificial Intelligence (AI)Skills: Machine Learning · Deep Learning · Artificial Intelligence (AI)
Hollywood Writers, AI and NFT TechnologyHollywood Writers, AI and NFT Technology

Associated with CloprNFTAssociated with CloprNFT
I led the development of CloprNFT as the product owner. Our groundbreaking technology seamlessly integrates stories penned by esteemed Hollywood writers with our cutting-edge in-house AI. Notably, each story is deliberately left open-ended for heightened engagement. What sets us apart is our automated customization process: utilizing computer vision and metadata available on the blockchain, each story is uniquely tailored to the image (NFT) triggering it.I led the development of CloprNFT as the product owner. Our groundbreaking technology seamlessly integrates stories penned by esteemed Hollywood writers with our cutting-edge in-house AI. Notably, each story is deliberately left open-ended for heightened engagement. What sets us apart is our automated customization process: utilizing computer vision and metadata available on the blockchain, each story is uniquely tailored to the image (NFT) triggering it.
Skills: Project Management · Machine Learning · Blockchain · Artificial Intelligence (AI)Skills: Project Management · Machine Learning · Blockchain · Artificial Intelligence (AI)

Clopr - Potion makers of web3Clopr - Potion makers of web3

Clopr LingoClopr Lingo

CloprClopr
Supervised Machine Learning (Classification task) - Conversion Rate PredictionSupervised Machine Learning (Classification task) - Conversion Rate Prediction

Associated with Jedha BootcampAssociated with Jedha Bootcamp
For this project, I embarked on a challenging endeavor in supervised machine learning known as the Conversion Rate Challenge. The goal? To predict whether a user visiting the Data Science Weekly website would subscribe to the newsletter based on various user parameters.

To tackle this challenge effectively, I followed a structured approach:

1. Exploratory Data Analysis (EDA) and Preprocessing: I delved into the provided dataset, uncovering insights and patterns to guide my model development. Understanding the data's nuances was crucial for preprocessing and feature engineering.

2. Baseline Model Training: With data_train.csv in hand, I trained a baseline model, setting the foundation for subsequent improvements. This initial model provided a benchmark against which I could measure enhancements.

3. Model Enhancement: Armed with insights from EDA, I embarked on enhancing my model's performance. This involved feature engineering, selecting pertinent features, experimenting with various algorithms, and fine-tuning hyperparameters to optimize the model's F1-score.

4. Prediction and Submission: Once satisfied with my model's performance, I utilized data_test.csv to make predictions. These predictions were then compiled into a submission file for evaluation, contributing to the ongoing leaderboard rankings.

5. Analysis and Recommendations: Beyond model performance, I delved into analyzing the parameters of my best model. This analysis aimed to unearth actionable insights for improving the newsletter's conversion rate. I provided recommendations based on my findings, potentially identifying new strategies to enhance user engagement.

Throughout this project, I adhered to the outlined goals:

- Conducting thorough EDA and preprocessing.
- Training and evaluating models for conversion prediction.
- Making submissions to the leaderboard.
- Analyzing model parameters and offering actionable recommendations.

Find my code in my GitHub.For this project, I embarked on a challenging endeavor in supervised machine learning known as the Conversion Rate Challenge. The goal? To predict whether a user visiting the Data Science Weekly website would subscribe to the newsletter based on various user parameters. To tackle this challenge effectively, I followed a structured approach: 1. Exploratory Data Analysis (EDA) and Preprocessing: I delved into the provided dataset, uncovering insights and patterns to guide my model development. Understanding the data's nuances was crucial for preprocessing and feature engineering. 2. Baseline Model Training: With data_train.csv in hand, I trained a baseline model, setting the foundation for subsequent improvements. This initial model provided a benchmark against which I could measure enhancements. 3. Model Enhancement: Armed with insights from EDA, I embarked on enhancing my model's performance. This involved feature engineering, selecting pertinent features, experimenting with various algorithms, and fine-tuning hyperparameters to optimize the model's F1-score. 4. Prediction and Submission: Once satisfied with my model's performance, I utilized data_test.csv to make predictions. These predictions were then compiled into a submission file for evaluation, contributing to the ongoing leaderboard rankings. 5. Analysis and Recommendations: Beyond model performance, I delved into analyzing the parameters of my best model. This analysis aimed to unearth actionable insights for improving the newsletter's conversion rate. I provided recommendations based on my findings, potentially identifying new strategies to enhance user engagement. Throughout this project, I adhered to the outlined goals: - Conducting thorough EDA and preprocessing. - Training and evaluating models for conversion prediction. - Making submissions to the leaderboard. - Analyzing model parameters and offering actionable recommendations. Find my code in my GitHub.
Skills: Machine LearningSkills: Machine Learning
Supervised Machine Learning (Regression task) - Weekly sales predictionSupervised Machine Learning (Regression task) - Weekly sales prediction

Associated with Jedha BootcampAssociated with Jedha Bootcamp
I was tasked with spearheading a project for Walmart's marketing service, aiming to construct a robust machine learning model capable of accurately estimating weekly sales in their stores. This endeavor was crucial for Walmart to comprehend the intricate relationship between sales and various economic indicators, thereby aiding in the formulation of informed marketing strategies.

The project was meticulously structured into three pivotal phases:

Part 1: I delved into extensive exploratory data analysis (EDA) and preprocessing to whip the dataset into shape for machine learning. This involved creating visualizations, computing

statistics, and implementing necessary transformations such as handling missing values, feature engineering from the Date column, and addressing outliers.

Part 2: The cornerstone of this phase was training a linear regression model, serving as our baseline. Following rigorous training, I meticulously assessed its performance on both the training and test sets. Moreover, delving into the model's coefficients provided valuable insights into the features driving sales predictions.

Part 3: To mitigate overfitting, I ventured into the realm of regularized regression models. Leveraging scikit-learn's Ridge and Lasso models, I aimed to strike a delicate balance between model complexity and generalization performance. Additionally, fine-tuning the regularization strength using GridSearchCV ensured optimal model performance.

Throughout the project, I adhered to Walmart's custom dataset, meticulously navigating through each step outlined in the project's scope and the code can be found in my GitHub.I was tasked with spearheading a project for Walmart's marketing service, aiming to construct a robust machine learning model capable of accurately estimating weekly sales in their stores. This endeavor was crucial for Walmart to comprehend the intricate relationship between sales and various economic indicators, thereby aiding in the formulation of informed marketing strategies. The project was meticulously structured into three pivotal phases: Part 1: I delved into extensive exploratory data analysis (EDA) and preprocessing to whip the dataset into shape for machine learning. This involved creating visualizations, computing statistics, and implementing necessary transformations such as handling missing values, feature engineering from the Date column, and addressing outliers. Part 2: The cornerstone of this phase was training a linear regression model, serving as our baseline. Following rigorous training, I meticulously assessed its performance on both the training and test sets. Moreover, delving into the model's coefficients provided valuable insights into the features driving sales predictions. Part 3: To mitigate overfitting, I ventured into the realm of regularized regression models. Leveraging scikit-learn's Ridge and Lasso models, I aimed to strike a delicate balance between model complexity and generalization performance. Additionally, fine-tuning the regularization strength using GridSearchCV ensured optimal model performance. Throughout the project, I adhered to Walmart's custom dataset, meticulously navigating through each step outlined in the project's scope and the code can be found in my GitHub.

Skills: Machine LearningSkills: Machine Learning
Unsupervised ML (DBSCAN + Truncated SVD) - Deploying a recommender system & Topic ExtractionUnsupervised ML (DBSCAN + Truncated SVD) - Deploying a recommender system & Topic Extraction

Associated with Jedha BootcampAssociated with Jedha Bootcamp
For the North Face ecommerce project, I embarked on utilizing unsupervised machine learning techniques to enhance online sales through the company's website. The primary focus was on implementing two major solutions to boost conversion rates.

Firstly, I deployed a recommender system aimed at suggesting additional products to users based on their interests. This involved creating a "you might also be interested by these products..." section on each product page, enhancing the user experience and potentially increasing sales.

Secondly, I endeavored to improve the structure of the product catalog by employing topic extraction techniques. The goal was to challenge existing categorizations and identify new categories that could enhance website navigation.

Breaking down the project into three main steps, I began by identifying groups of products with similar descriptions. This involved preprocessing textual data to clean the corpus, encoding the texts with TF-IDF transformation, and training a clustering model using DBSCAN on the TF-IDF matrix.

Subsequently, I developed a recommender system by utilizing the cluster IDs obtained from the clustering model. This allowed me to suggest similar items to users based on their interests, enhancing the personalized shopping experience.

Finally, I delved into topic modeling using an LSA model to automatically extract latent topics from the product descriptions. This involved training a TruncatedSVD model on the corpus and analyzing the results through wordclouds to gain insights into the underlying themes present in the product descriptions.

In conclusion, the project aimed to leverage unsupervised machine learning techniques to optimize the online shopping experience for The North Face customers, potentially increasing conversion rates and overall sales on the company's website. Find the code in my GitHub.For the North Face ecommerce project, I embarked on utilizing unsupervised machine learning techniques to enhance online sales through the company's website. The primary focus was on implementing two major solutions to boost conversion rates. Firstly, I deployed a recommender system aimed at suggesting additional products to users based on their interests. This involved creating a "you might also be interested by these products..." section on each product page, enhancing the user experience and potentially increasing sales. Secondly, I endeavored to improve the structure of the product catalog by employing topic extraction techniques. The goal was to challenge existing categorizations and identify new categories that could enhance website navigation. Breaking down the project into three main steps, I began by identifying groups of products with similar descriptions. This involved preprocessing textual data to clean the corpus, encoding the texts with TF-IDF transformation, and training a clustering model using DBSCAN on the TF-IDF matrix. Subsequently, I developed a recommender system by utilizing the cluster IDs obtained from the clustering model. This allowed me to suggest similar items to users based on their interests, enhancing the personalized shopping experience. Finally, I delved into topic modeling using an LSA model to automatically extract latent topics from the product descriptions. This involved training a TruncatedSVD model on the corpus and analyzing the results through wordclouds to gain insights into the underlying themes present in the product descriptions. In conclusion, the project aimed to leverage unsupervised machine learning techniques to optimize the online shopping experience for The North Face customers, potentially increasing conversion rates and overall sales on the company's website. Find the code in my GitHub.
Skills: Machine Learning · Cluster AnalysisSkills: Machine Learning · Cluster Analysis
Unsupervised ML (KMeans vs DBSCAN) - Hot zone clustering for taxi pick upUnsupervised ML (KMeans vs DBSCAN) - Hot zone clustering for taxi pick up

Associated with Jedha BootcampAssociated with Jedha Bootcamp
I led a project at Uber focused on optimizing pickup locations through unsupervised machine-learning techniques. The challenge we aimed to address was the inconvenience users faced when drivers weren't readily available in their vicinity, leading to longer wait times or ride cancellations.

To tackle this, my team and I developed algorithms to identify hot-zones in major cities, starting with New York City as our initial test bed. Leveraging Uber's extensive trip data, we employed clustering techniques, particularly KMeans and DBScan, to group pickup locations into clusters, thereby pinpointing areas of high demand at various times of the day.

Our approach involved starting small, focusing on specific days and hours, before expanding to generalize our findings across different timeframes. Using Python libraries like Plotly, we visualized these hot-zones on interactive maps, offering a clear understanding of where drivers should be positioned for optimal service.

Find the code in my GitHub.I led a project at Uber focused on optimizing pickup locations through unsupervised machine-learning techniques. The challenge we aimed to address was the inconvenience users faced when drivers weren't readily available in their vicinity, leading to longer wait times or ride cancellations. To tackle this, my team and I developed algorithms to identify hot-zones in major cities, starting with New York City as our initial test bed. Leveraging Uber's extensive trip data, we employed clustering techniques, particularly KMeans and DBScan, to group pickup locations into clusters, thereby pinpointing areas of high demand at various times of the day. Our approach involved starting small, focusing on specific days and hours, before expanding to generalize our findings across different timeframes. Using Python libraries like Plotly, we visualized these hot-zones on interactive maps, offering a clear understanding of where drivers should be positioned for optimal service. Find the code in my GitHub.
Skills: Machine Learning · Cluster Analysis