UBS

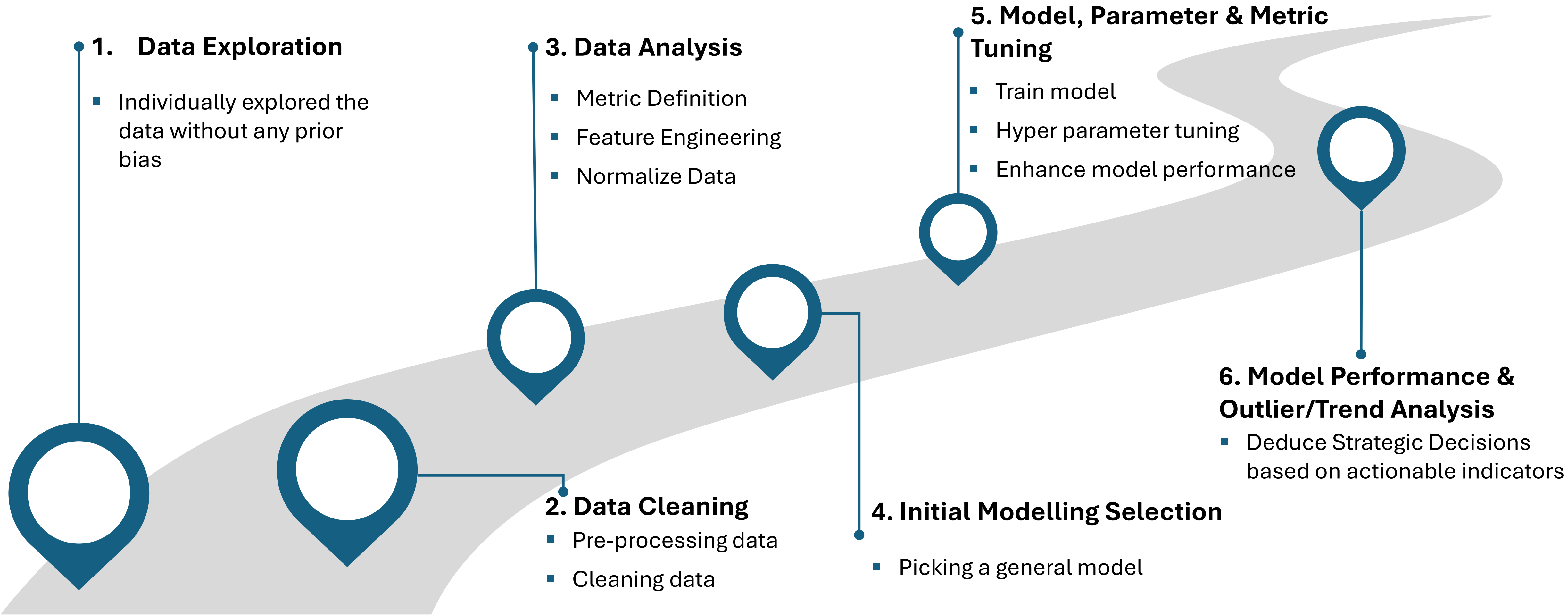
DATATHON 2024 - UBS CHALLENGE

Hackermen Team:

- Ishaan Bhondele
- Virgillio Strozzi
- Andrea Ghirlanda
- Alexander Lerch

April 28

Roadmap



1. Data Exploration

- Individually explored the data without any prior bias

2. Data Cleaning

- Pre-processing data
- Cleaning data

3. Data Analysis

- Metric Definition
- Feature Engineering
- Normalize Data

4. Initial Modelling Selection

- Picking a general model

5. Model, Parameter & Metric Tuning

- Train model
- Hyper parameter tuning
- Enhance model performance

6. Model Performance & Outlier/Trend Analysis

- Deduce Strategic Decisions based on actionable indicators

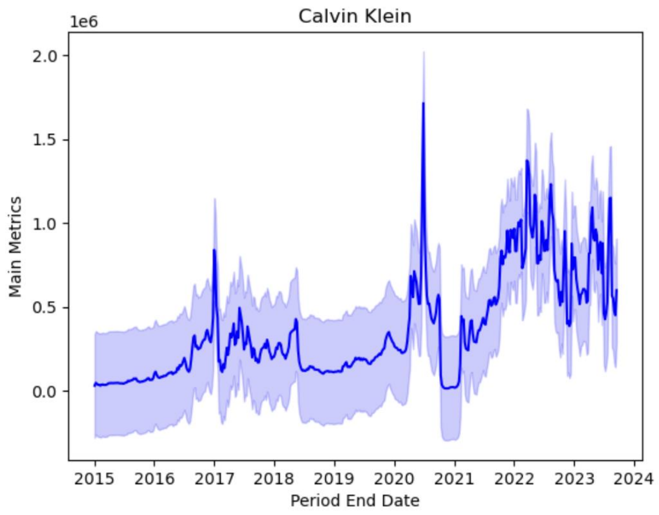
Data Exploration

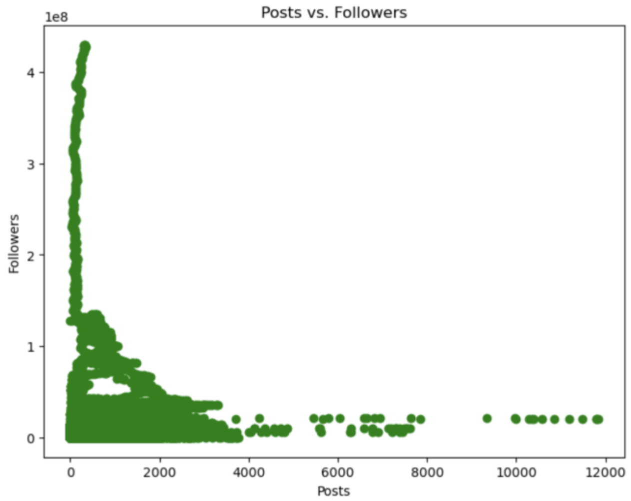
Our Goal is to get familiar with the dataset by

Analyzing the unique values in each column (business_entity_doing_business_as_name, primary_exchange_name, etc.)

Looking at the distribution (amount of data available) on a yearly basis.

Finding Potential columns that could be used for features





Data Exploration

Data Cleaning

Data Analysis

Initial Modelling Selection

Model, Parameter & Metric Tuning

Model Performance & Outlier/Trend Analysis

Data Cleaning

Our goal is to preprocess and clean the dataset. The key steps are:

- Investigating the missing values and checking for best replacement
- Dropping NaN and duplicates
- Removing unwanted columns
- Implementing Timedelta instead of Absolute datetime (to see if there any possible benefits)

Data Exploration

Data Cleaning

Data Analysis

Initial Modelling Selection

Model, Parameter & Metric Tuning

Model Performance & Outlier/Trend Analysis

Data Analysis

Our goal is to define Features/Metrics that can be passed into the model:

We define 11 metrics that are then passed to the model

The metrics are defined based on interaction (likes & comments), followers, posts (pictures & videos) and difference in likes.

Data Exploration

Data Cleaning

Data Analysis

Initial Modelling Selection

Model, Parameter & Metric Tuning

Model Performance & Outlier/Trend Analysis

Initial Model Selection

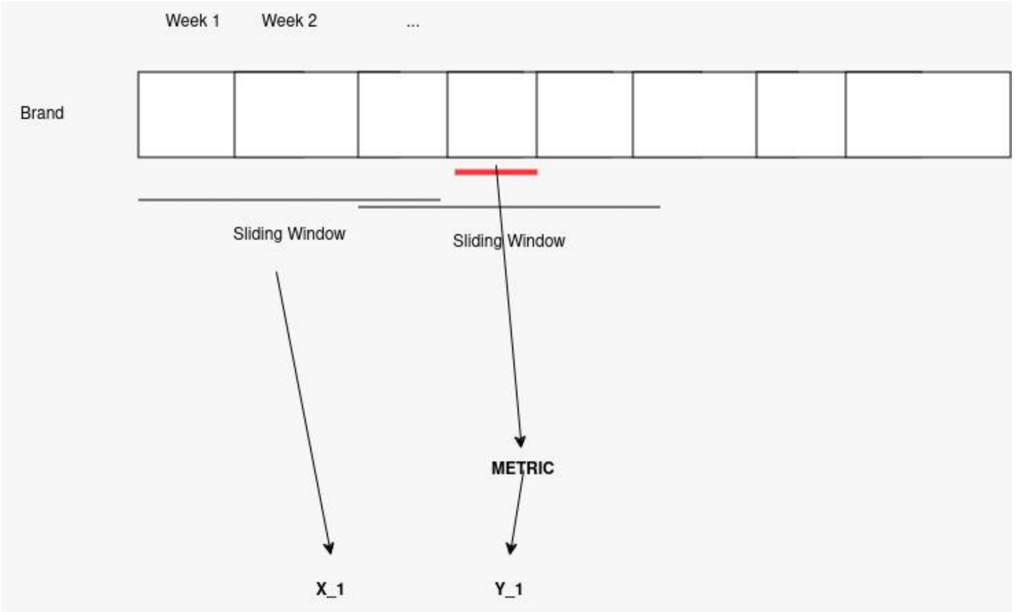
Key Metric:

Capture local-growth and defined as: interactions/posts, where interactions is a weighted sum over likes, comments and followers per week

Type of Model:

We try two models that are a simple LSTM implementation and a ConvRNN

Classification is now over a time-period



This is good because:

- The models captures dependencies inside the window of week to predict the future metric value.
- Hence, we opt for two models which have the right bias to capture this.
- We use as a Loss a simple Mean Squared Error and we evaluate the prediction still with the Mean Squared Error

Data Exploration | Data Cleaning | Data Analysis | Initial Modelling Selection | Model, Parameter & Metric Tuning | Model Performance & Outlier/Trend Analysis

Model, Parameter & Metric Tuning

Assumptions:

Don’t need lots of consecutive weeks to detect a positive trend.
We train over data before 2022 first and then finetune on data after 2022.

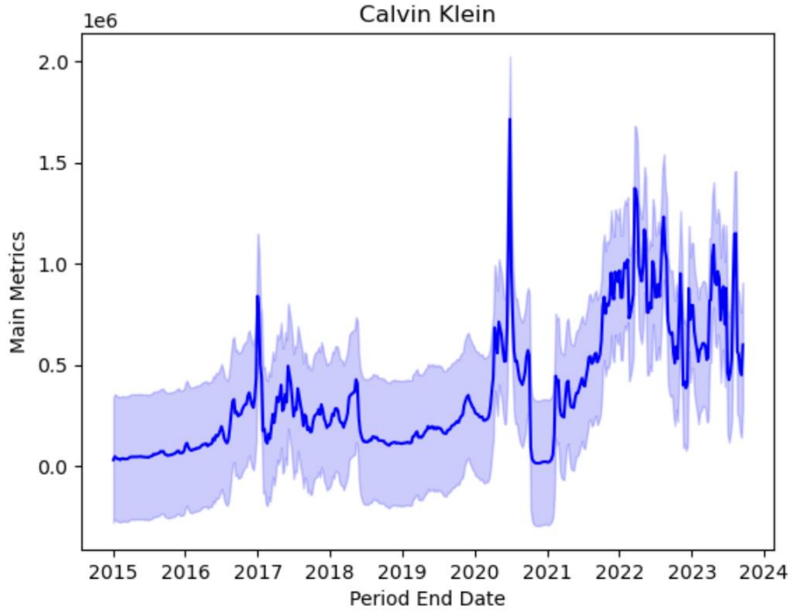
Growth locally (Brand specific)

Data Exploration | Data Cleaning | Data Analysis | Initial Modelling Selection | **Model, Parameter & Metric Tuning** | Model Performance & Outlier/Trend Analysis

Model Performance & Outlier/Trend Analysis

The prediction can be used inside a Test to check if the brand is an outlier compared to its previous trend. To do that we proceed as follow:

- We evaluate the metric over the window_length to generate an *avg_growth* in the past period. Moreover we compute the *std_dev_growth* from the metric growth for each week inside our window_length. Our model is then making the prediction *growth* of the metric for the future in the next K=1 weeks
- We then check whether *growth* - *avg_growth* > *zstd_dev_growth*, z tunable (ex. z=2 means in 95.47% positive outlier) to detect whether we have a **POSITIVE OUTLIER**



Data Exploration

Data Cleaning

Data Analysis

Initial Modelling Selection

Model, Parameter & Metric Tuning

Model Performance & Outlier/Trend Analysis

UBS

Lessons learned & Open Points

Quickly drop irrelevant features & data records

Get simple model running first!

And then improve iteratively

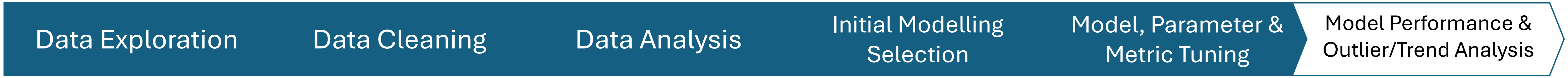
Define target and key metrics

And then improve iteratively

Develop a model with:

Significant deviations from observed trends → Define “interesting” deviation

To highlight noteworthy brands based on the provided dataset


Data Exploration

Data Cleaning

Data Analysis

Initial Modelling Selection

Model, Parameter & Metric Tuning

Model Performance & Outlier/Trend Analysis