



dP | delphi  
Project

Our **MISSION** is to create a tool that reliably captures the daily sentiment in a given group of people regarding a specific topic

**In short:** Create more reliable polls!

**Project 1:** Predict the German Elections 2021 if they happened today

## Current problems in traditional polls and surveys

01  
Time lag

02  
„Shy Trump“ Theory

03  
Expensive

We analyse millions of tweets and poll data for each party ...

Our solution

1.3m

&

...and produce a predictive outcome on the German election.

**CDU**

**23%**

**SPD**

**20%**

**Greens**

**18%**

# How it works



We automatically **fetch** data from the Twitter API every day...



**analyse** the sentiment of each Tweet and **create** model features ...



**predict** the outcome of the election if it happened today ...



and **present** our findings in a dashboard for customers.

# How it works



We automatically fetch data from the Twitter API...



analyse the sentiment of each Tweet and create model features ...

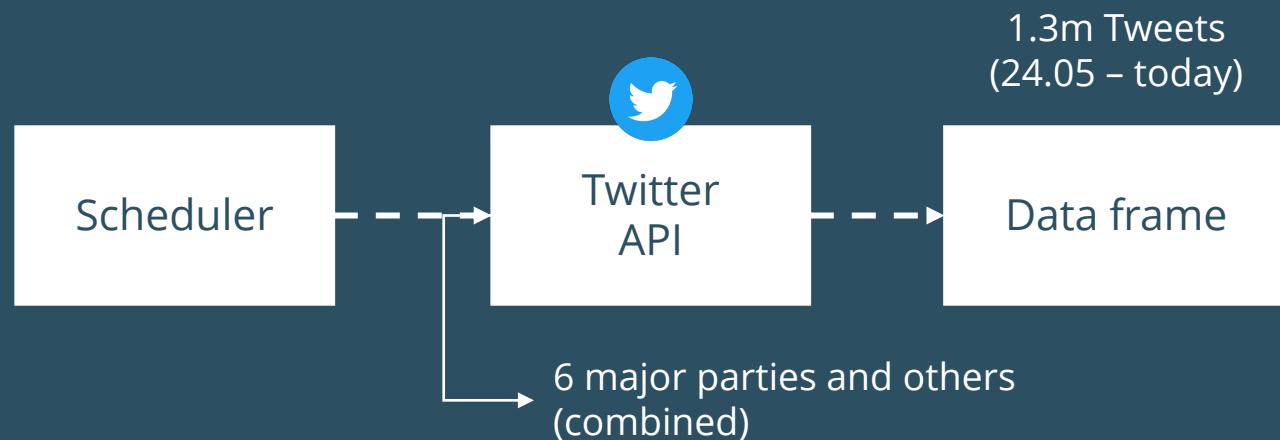


predict the outcome of the election if it happened today ...



and present our findings in a dashboard for customers.

## Workflow



## Deep-Dive

- At midnight the Google scheduler calls our API with a post request
- Twitter is searched with an academic API endpoint for seven predefined keyword (last 24 hours)
- This results in one data frame containing all tweets matching our search string labelled with their according party

# How it works



We automatically fetch data from the Twitter API every day...



analyse the sentiment of each Tweet and create model features ...

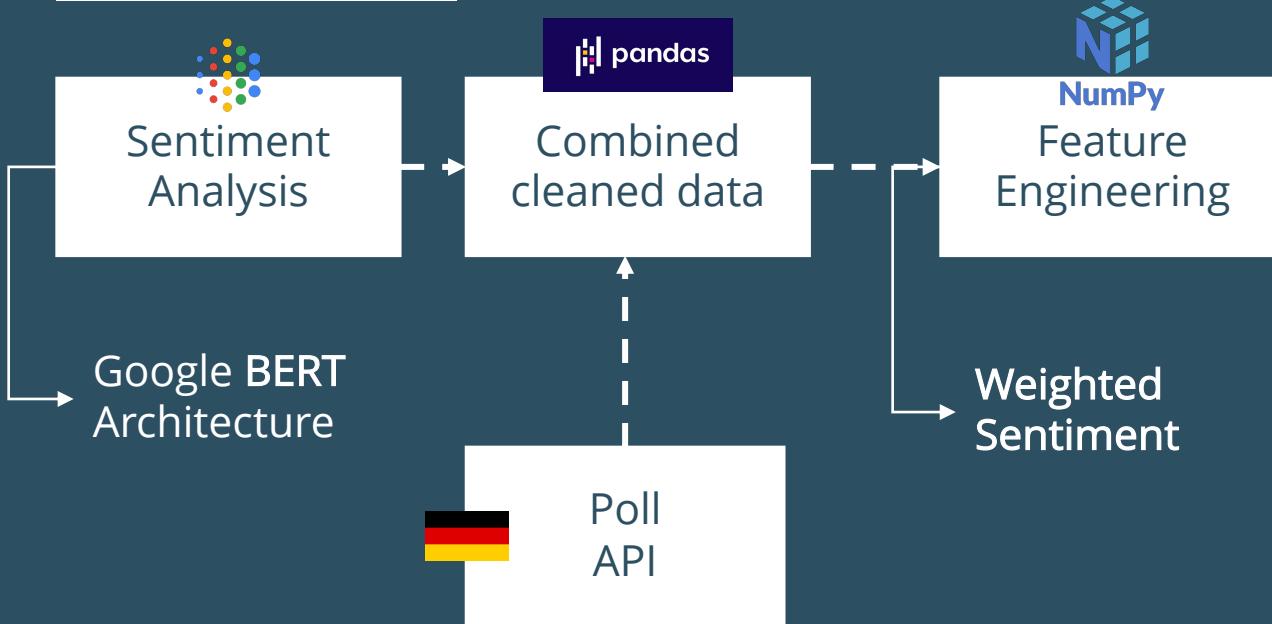


predict the outcome of the election if it happened today ...

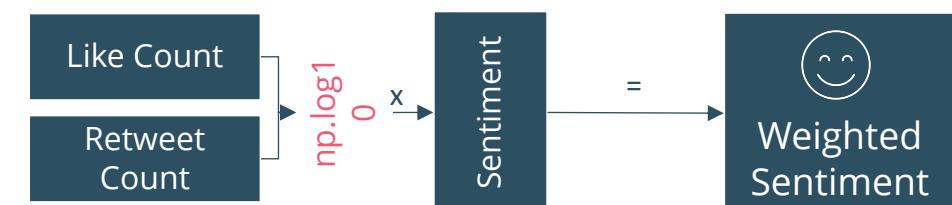


and present our findings in a dashboard for customers.

## Workflow



## Deep-Dive



- NLP Model classifies each tweet in one of three sentiment categories
- Poll data is fetched for the last 3 months
- Combined cleaned data frame groups the 1.3m tweets per day (incl. sentiment) and political party

# How it works



We automatically fetch data from the Twitter API every day...



analyse the sentiment of each Tweet and **create** model features ...

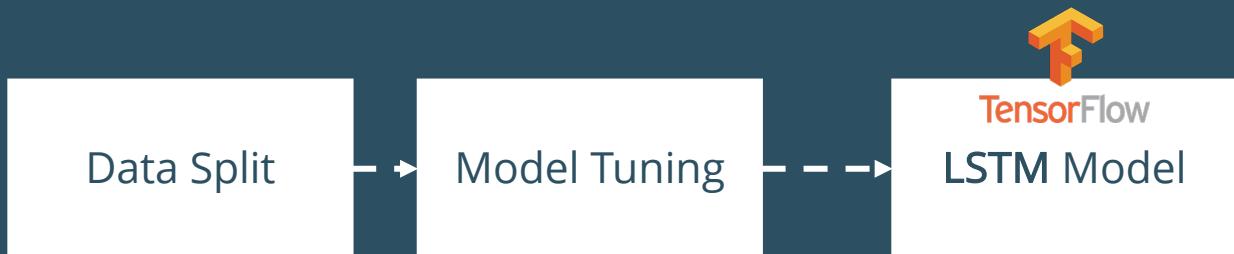


**predict** the outcome of the election if it happened today ...



and **present** our findings in a dashboard for customers.

## Workflow



## Deep-Dive

- Creating 88 subsequences (4 days each)
- The best features are identified (most impact on the RNN model) and hyper-parameters are tuned
- Using the Long short-term memory (LSTM) model in order to achieve best performance

# How it works



We automatically fetch data from the Twitter API every day...



analyse the sentiment of each Tweet and **create** model features ...



**predict** the outcome of the election if it happened today ...



and **present** our findings in a dashboard for customers.

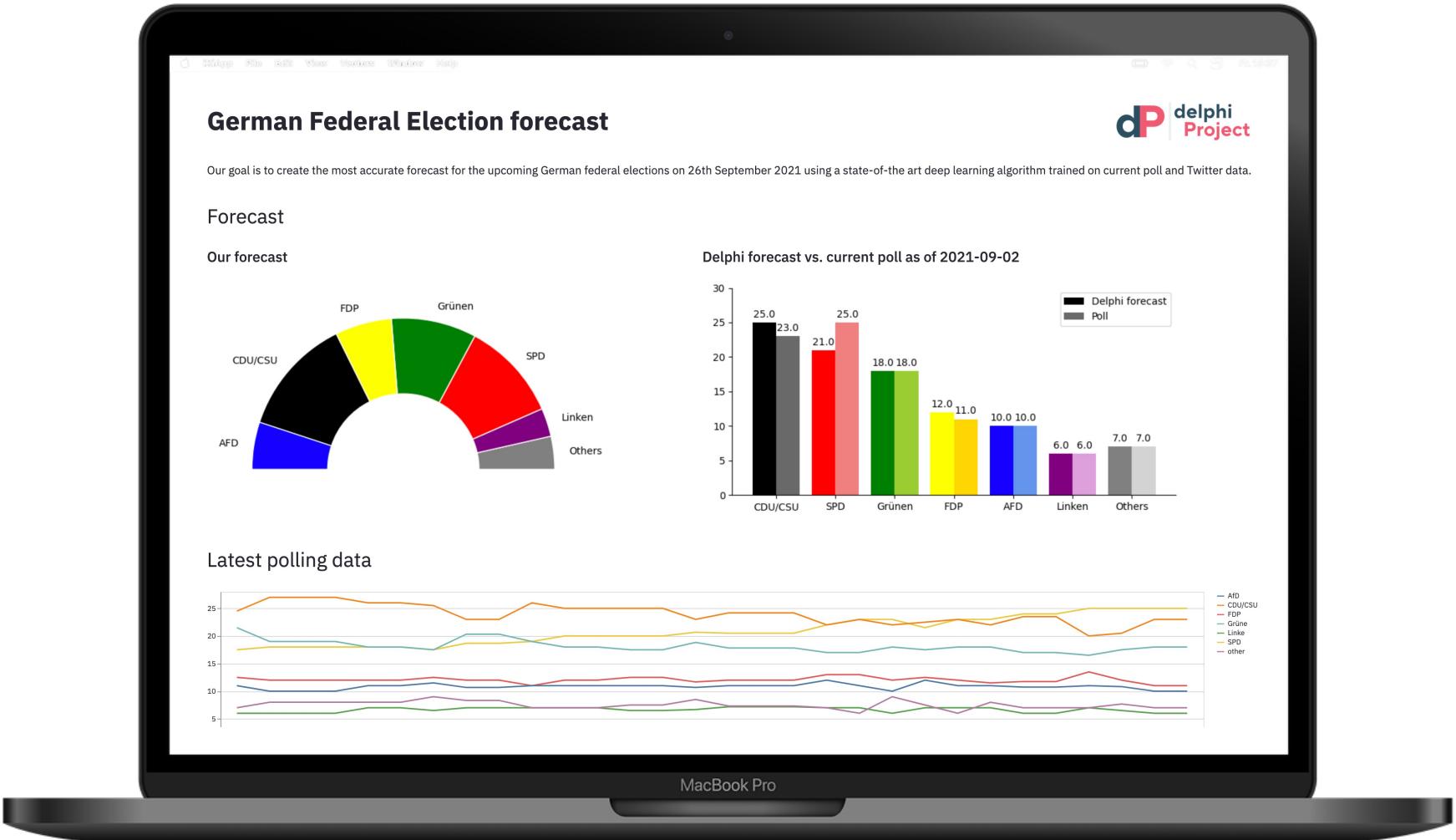
## Workflow



## Deep-Dive

- Retrieve data during feature creation process and upload to Google Cloud Storage
- Deployed website on Heroku retrieves the data from GCP every time the site is loaded
- Charts and Tweet data will be updated according to users selection

# Our final product



Start Product Demo

The team  
that made it  
happen

# The Team



Luca

The Git Master



Nicolas

The OTHER(S) Master



Philipp

XYZ



Finn

The Notebook Master

Focus Areas



Google Cloud



Background



le wagon



le wagon



le wagon



le wagon

# Appendix



# Current problems in traditional polls and surveys

## 01 Time lag

- Polls have to be carefully planned in advance in order to maximize the representative nature of the poll
- Polls are usually being conducted over several days and locations, which takes time
- Impossible to capture daily sentiment changes (e.g. a scandal)

## 02 „Shy Trump“ Theory

- In the 2016 US Presidential election, a phenomenon occurred in which voters would be too „shy“ to tell pollsters that they would eventually vote for Trump
- This leads to an underrepresentation for extreme/controversy candidates or for groups that generally mistrust public institutions

## 03 Expensive

- Traditional polls are expensive to carry out, as they still include many unautomated tasks
- If a customer wants to make a survey/poll on a specific topic, they usually have high upfront costs

## We produce a reliable sentiment on a given population...

### Our solution

#### Updated daily

- All available poll and social media data is fetched on a daily basis
- Our proprietary tool is based on millions of individual data points, that are categorized into different "data groups" (i.e. parties participating in the #BTW21)
- Uses SM data to avoid the "shy Trump" problem

#### Sentiment Analysis

- Each individual data point (i.e. Tweet or social media post) is analysed with a Natural Language Processing (NLP) model in order to determine the overall sentiment of that data point
- Afterwards multiple new features are generated that are used in the deep learning model (e.g. weighted sentiment)

#### Deep Learning Model

- Our Recurrent Neural Network (RNN) model (LSTM) analyses the features for each party on a given day and learns the intricacies of the poll and social media developments
- The output of the model is a prediction of the % the different parties would get on that specific day

&

**...and enable customers to make data-driven decisions.**

#### Easy

- We offer our customers a customer centric user experience
- All of the available insights are presented in an easy and intuitive dashboard

1.3m

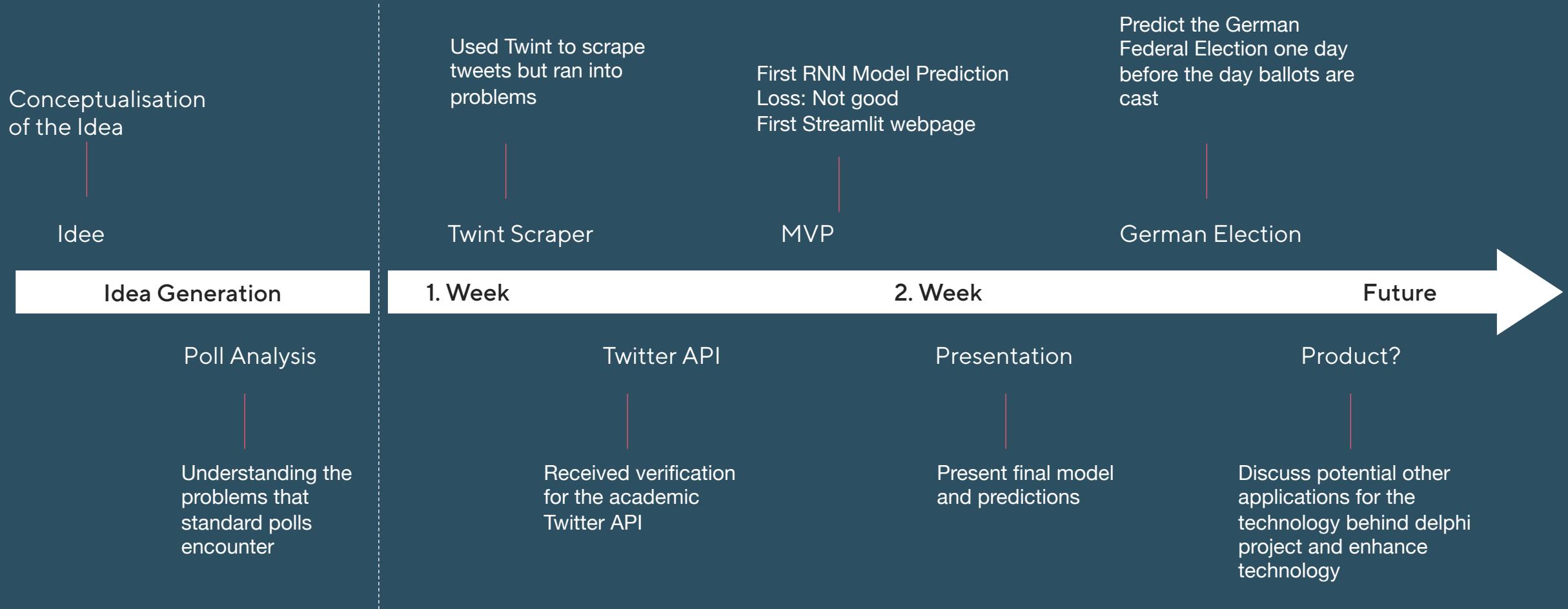
Tweets<sup>1</sup>  
Analyzed

For the German  
elections we analyzed  
1.3m tweets over six  
weeks...

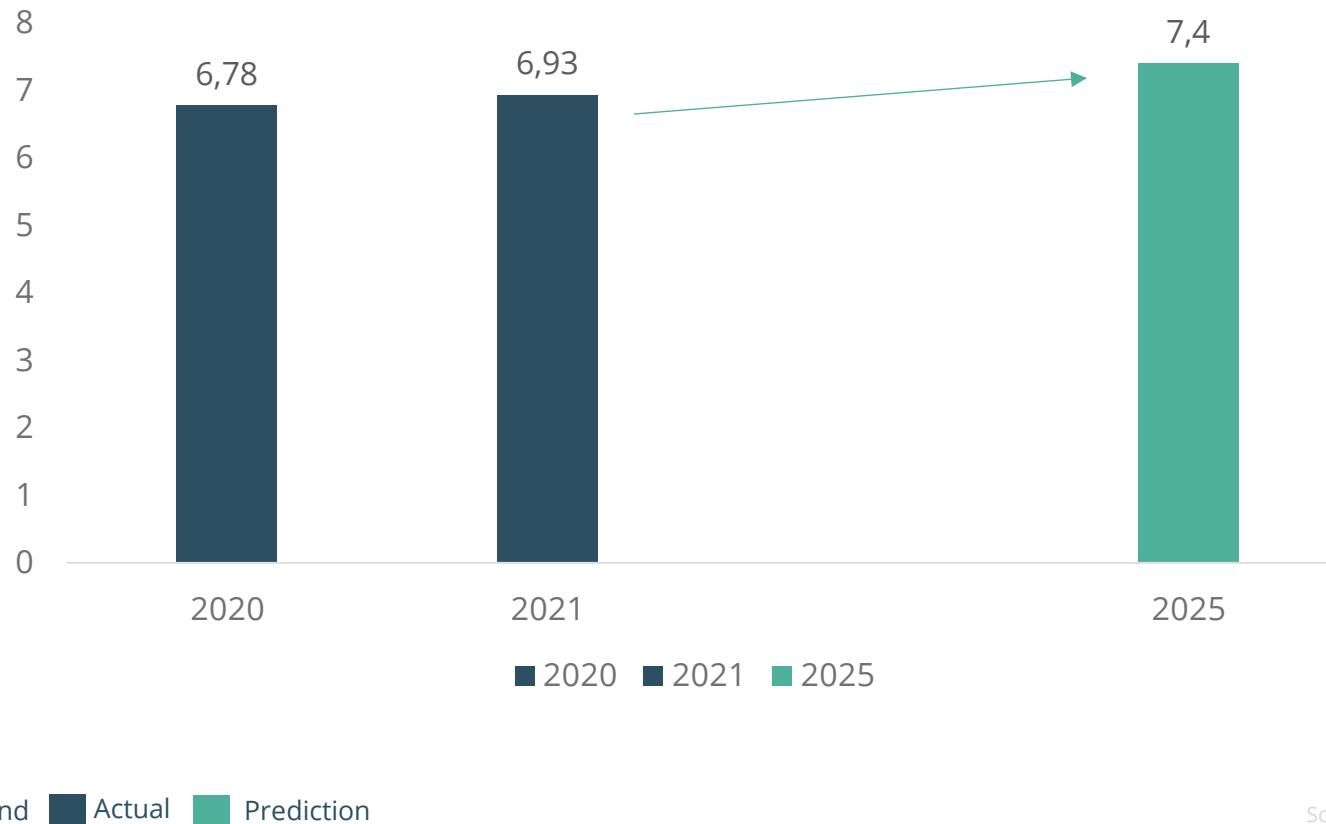


<sup>1</sup>Tweets = Only tweets are included that can be clearly linked to one specific party

# Roadmap



## Global industry revenue of „Public Opinion and Election Polling” (in Billion USD)



Source:  
Global News Wire

\$7bn  
USD

Global market size is around USD 7bn. A minor market share would still result in a big opportunity.

E.G: TAM = 700M