Machine Learning Canvas

Product:

Social Media Trend Analyse

Sofie Pischl

Date:

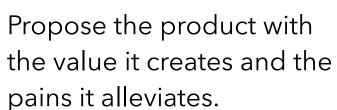
05.04.2025

Background

Describe the customer's goals and pains.

Verlage, Autoren und Content-Creator haben Schwierigkeiten, frühzeitig relevante Trends zu erkennen. Die manuelle Recherche ist sehr zeitaufwändig und unzuverlässig (häufig beeinflusst von eigenen Interessen und Nutzungsmustern).

Value proposition



Das System analysiert Social-Media-Daten (z.B. Instagram, Reddit, TikTok) automatisch und erkennt aufkommende Themen.

Dadurch können Verlage zielgerichteter auf Marktbewegungen reagieren – ohne aufwändige Recherchen oder teure Trendberater.

So können passende Buchideen oder anderer Content entwickelt werden und der Marktanteil gesteigert werden.

key objectives that need to

- Sentiment-Analyse zu bestimmten Themen

Instagram, Reddit) - Extraktion von Trending Topics

Solution

Define the solution, including features, integration, constraints and what's out-of-scope

- Daten aus APIs (Instagram, Reddit, etc.) sammelnNLP-Pipeline mit Preprocessing + Topic Modeling via
- Sentiment-Analyse mit vortrainiertem Modell von
- Visualisierung mit Streamlit (schnell, einfach)
- Tägliches Batch-Update, keine Realtime-Verarbeitung - Deployment lokal oder via kostenlosem Hosting (z.B. Streamlit Cloud)

Data

Identify the training and production data sources, as well as the labeling process and decisions.

- Quellen: Twitter API, Reddit API, ggf. TikTok
- Labeling: Unsupervised (Topic Modeling), optional Crowd-Labeling für Sentiment - Speicherung: Cloud-Datenbank (z.B. BigQuery,



Authors:

Modeling

List the iterative approach to model our task.

- Embedding: Sentence-BERT
- (all-MiniLM-L6-v2) via sentence-transformers
- Clustering: HDBSCAN via BERTopic Dimension Reduktion: UMAP
- Sentiment: HuggingFace-Modell (z.B. cardiffnlp/twitter-roberta-base-sentiment)



Feedback



Version:

Outline sources of feedback from our system to use for iteration.

- Eigene Bewertung (macht das Thema
- Vergleich mit Google Trends - Optionales User-Feedback über
- Formular oder CSV-Review

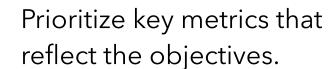
Project



Define the required team members, deliverables and projected timelines.

Woche 1 – Setup, APIs, Datenerhebung Woche 2 – Preprocessing + Sentiment Woche 3 – Topic Modeling & Visualisierung Woche 4 – Dashboard & Evaluation optional: Woche 5-6 für Feinschliff, CI/CD, Hosting

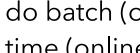
Metrics



- Precision / Recall der Trendklassifikation
- Zeit bis zur Trend-Erkennung
- Sentimentverteilung zu bestimmten Themen

Decide whether we want to do batch (offline) or realtime (online) inference.

> Batchweise Analyse, z.B. einmal täglich. Kein Realtime, kein Docker nötig. Automatisierung über einfaches Skript + Scheduler (z.B. Python



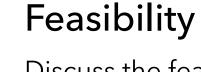
Inference

schedule, oder später: Airflow/CRON).



Breakdown the product into be delivered.

- Daten von Social-Media-Plattformen extrahieren (z.B.
- Webinterface mit einfacher Visualisierung bereitstellen



Discuss the feasibility of the solution and if we have the required resources.

- Alles mit Open-Source-Tools realisierbar - Keine GPU zwingend erforderlich (BERTopic
- kann auch mit Sentence-BERT + UMAP laufen) - Datenquellen mit kostenloser Nutzung (ggf. Ratenlimit beachten)
- Alles in Python umsetzbar (du brauchst nur ein Notebook + Editor + GitHub)



Design offline and online evaluation criteria.

- Offline: Visualisierung der Topics und Sentiments, Plausibilitätsprüfung durch Nutzer
- Kriterium: Erkennbare, nachvollziehbare Trends mit hoher Relevanz





Machine learning canvas from Made With ML by Goku Mohandas

License: <u>CC BY-SA 4.0</u>