

Machine Learning Canvas

Product:

Social Media Trend Analyse

Authors:













Sofie Pischl

Date:

05.04.2025

Version:

1

<div>Background</div> <div></div> <div>Describe the customer’s goals and pains.</div> <div><p>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).</p></div>	<div>Solution</div> <div></div> <div>Define the solution, including features, integration, constraints and what's out-of-scope</div> <div><ul style="list-style-type: none">- Daten aus APIs (Instagram, Reddit, etc.) sammeln- NLP-Pipeline mit Preprocessing + Topic Modeling via BERTopic- Sentiment-Analyse mit vortrainiertem Modell von HuggingFace- Visualisierung mit Streamlit (schnell, einfach)- Tägliches Batch-Update, keine Realtime-Verarbeitung- Deployment lokal oder via kostenlosem Hosting (z.B. Streamlit Cloud)</div>	<div>Data</div> <div></div> <div>Identify the training and production data sources, as well as the labeling process and decisions.</div> <div><ul style="list-style-type: none">- Quellen: Twitter API, Reddit API, ggf. TikTok Scraper- Labeling: Unsupervised (Topic Modeling), optional Crowd-Labeling für Sentiment- Speicherung: Cloud-Datenbank (z.B. BigQuery, MongoDB)</div>	<div>Modeling</div> <div></div> <div>List the iterative approach to model our task.</div> <div><ul style="list-style-type: none">- 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)</div>	<div>Feedback</div> <div></div> <div>Outline sources of feedback from our system to use for iteration.</div> <div><ul style="list-style-type: none">- Eigene Bewertung (macht das Thema Sinn?)- Vergleich mit Google Trends- Optionales User-Feedback über Formular oder CSV-Review</div>
<div>Value proposition</div> <div></div> <div>Propose the product with the value it creates and the pains it alleviates.</div> <div><p>Das System analysiert Social-Media-Daten (z.B. Instagram, Reddit, TikTok) automatisch und erkennt aufkommende Themen.</p><p>Dadurch können Verlage zielgerichteter auf Marktbewegungen reagieren – ohne aufwändige Recherchen oder teure Trendberater.</p><p>So können passende Buchideen oder anderer Content entwickelt werden und der Marktanteil gesteigert werden.</p></div>		<div>Metrics</div> <div></div> <div>Prioritize key metrics that reflect the objectives.</div> <div><ul style="list-style-type: none">- Precision / Recall der Trendklassifikation- Zeit bis zur Trend-Erkennung- Sentimentverteilung zu bestimmten Themen</div>		<div>Project</div> <div></div> <div>Define the required team members, deliverables and projected timelines.</div> <div><p>Zeitplan:</p><p>Woche 1 – Setup, APIs, Datenerhebung</p><p>Woche 2 – Preprocessing + Sentiment</p><p>Woche 3 – Topic Modeling & Visualisierung</p><p>Woche 4 – Dashboard & Evaluation</p><p>optional: Woche 5–6 für Feinschliff, CI/CD, Hosting</p></div>
<div>Objectives</div> <div></div> <div>Breakdown the product into key objectives that need to be delivered.</div> <div><ul style="list-style-type: none">- Daten von Social-Media-Plattformen extrahieren (z.B. Instagram, Reddit)- Extraktion von Trending Topics- Sentiment-Analyse zu bestimmten Themen- Webinterface mit einfacher Visualisierung bereitstellen</div>	<div>Feasibility</div> <div></div> <div>Discuss the feasibility of the solution and if we have the required resources.</div> <div><ul style="list-style-type: none">- 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)</div>	<div>Evaluation</div> <div></div> <div>Design offline and online evaluation criteria.</div> <div><ul style="list-style-type: none">- Offline: Visualisierung der Topics und Sentiments, Plausibilitätsprüfung durch Nutzer- Kriterium: Erkennbare, nachvollziehbare Trends mit hoher Relevanz- Iteration basierend auf eigenem Eindruck oder externen Benchmarks</div>	<div>Inference</div> <div></div> <div>Decide whether we want to do batch (offline) or real-time (online) inference.</div> <div><p>Batchweise Analyse, z.B. einmal täglich. Kein Realtime, kein Docker nötig. Automatisierung über einfaches Skript + Scheduler (z.B. Python schedule, oder später: Airflow/CRON).</p></div>	

