**Titel:**

**Analyse von Musikgenres anhand von YouTube-Kommentaren mit VADER und LIWC**

**Ziel der Arbeit:**

Die Arbeit untersucht, wie sich Stimmungen und Polarität von Kommentaren auf YouTube zwischen verschiedenen Musikgenres unterscheiden lassen. Hierzu werden VADER-Sentiment-Scores und LIWC-Merkmale verwendet. Die Analyse soll systematisch auf mehreren Ebenen stattfinden: auf Song-, Künstler- und Genre-Ebene.

**Vorgehensweise & ToDos (geordnet):**

**1. Auswahl der Analysezeiträume**

* **Ziel:** Zeitkomponente minimieren, um externe Einflussfaktoren (z. B. COVID) zu vermeiden.
* **Vorgabe:** Nur Kommentare aus einem ähnlichen Zeitraum verwenden (z. B. **2023–2025**).
* **Maßnahme:** Kommentare beim Scraping zeitlich filtern.

**2. Genre-Definition & Auswahl relevanter Songs**

* Es werden **100 Mikrogenres** betrachtet.
* Für jedes Genre sollen bis zu **10 repräsentative Songs** bzw. Musikvideos identifiziert werden.
  + Beispiel:
    - Genre1 → Song1, Song2, …, Song10
    - Genre2 → Song1, Song2, …, Song10
* Wenn ein Song nur für ein Genre geeignet ist, wird er diesem exklusiv zugeordnet.
* Ansonsten inverse( dass wenn song 1 den höchsten wert für ein Genre hat dann wird dieser song für ein anderes Genre nicht mehr verwendet) finetune diese Logic

**3. Scraping & Preprocessing**

* **Neu-Scraping der YouTube-Kommentare** für die final ausgewählten Songs.
* Pro Genre sollen möglichst **gleich viele Kommentare** gesammelt werden.
* Vorverarbeitung der Kommentare (z. B. Sprache filtern, irrelevante Beiträge entfernen).

**4. Feature-Extraktion**

* Berechnung des **VADER-Compound-Scores** je Kommentar.
* Extraktion von **LIWC-Features** mit Python (z. B. gemäß [diesem Artikel](https://medium.com/bright-ai/nlp-feature-extraction-from-liwc-in-python-ded333bc97aa)).

**5. Analytische Auswertung (mehrstufig und vergleichend)**

* Vergleich von Sentiment- und LIWC-Scores:
  + **auf Song-Ebene**
  + **auf Künstler-Ebene**
  + **auf Genre-Ebene**
* Besondere Fragestellungen:
  + Welche Künstler polarisieren besonders?
  + Gibt es Genres mit überdurchschnittlich positiven oder negativen Kommentaren?
  + Korrelation von Song- und Künstler-Scores innerhalb eines Genres.

**6. Validierung**

* Sobald die Scraping- und Genre-Zuordnungslogik steht:
  + **E-Mail an Markus** senden zur Freigabe der finalen Liste an Videos/Songs.
  + Feedback abwarten und ggf. anpassen.

**Zeitplan (vorgeschlagen)**

| **Phase** | **Zeitraum** | **Ziel/Milestone** |
| --- | --- | --- |
| **1. Vorbereitungsphase** | 05.–10. August 2025 | Auswahl & Zuweisung von Songs pro Genre abschließen |
| **2. Scraping-Phase** | 11.–20. August 2025 | Scrape-Logik implementieren und Kommentare sammeln |
| **3. Preprocessing** | 21.–25. August 2025 | Kommentare filtern, normalisieren, Sprache erkennen |
| **4. Feature-Extraktion** | 26.–30. August 2025 | VADER & LIWC-Analyse vorbereiten |
| **5. Tiefenanalyse & Vergleich** | 01.–10. September 2025 | Vergleichende Auswertung auf Song-, Künstler-, Genre-Ebene |
| **6. Visualisierung & Auswertung** | 11.–15. September 2025 | Ergebnisse visuell aufbereiten |
| **7. Feinschliff & Schreiben** | 16.–30. September 2025 | Diskussion, Abstract, Einleitung & Fazit schreiben |
| **8. Abgabevorbereitung** | 01.–05. Oktober 2025 | Letzte Korrekturen, Zitierungen, Layout |