1. Kolom Data Set

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Nama | Kalimat | Tema\_postingan | ponpen | Aspek\_manual |

1. mengindifikasi aspek, kategorinya apa

|  |  |  |
| --- | --- | --- |
| No | Aspek Manual | Deskripsi |
| 1 | Kualitas Guru | Apresiasi terhadap guru, ustaz, pimpinan, atau tenaga pengajar. |
| 2 | Sosial | Terkait bantuan, kepedulian sosial, donasi, atau kegiatan sosial pondok. |
| 3 | Motivasi/Spiritual | Doa, harapan, semangat, dan ekspresi spiritual atau religius. |
| 4 | Kegiatan Pondok | Acara resmi, kunjungan, kegiatan internal pondok, rapat, atau seremonial. |
| 5 | Akademik | Ujian, wisuda, prestasi, hafalan, atau kegiatan pendidikan formal. |
| 6 | Umum | Komentar umum tanpa makna spesifik, seperti “MasyaAllah”, “Aamiin”, emoji. |
| 7 | Fasilitas | Komentar tentang bangunan, sarana, kamar, masjid, asrama, atau infrastruktur. |
| 8 | Lingkungan | Menyebut suasana pondok, alam sekitar, ketenangan, kebersihan, kenyamanan. |
| 9 | Pembinaan Karakter | Tentang akhlak, kedisiplinan, adab, tanggung jawab, atau nilai moral. |

1. Mencari aspek menggunakan LDA dari skearn atau library lain
2. Mengisi table human dan AI generate
3. melaraskan jumlah semua aspek dan sentiment dibantu sama ai untuk menggenerate(unique) dichat lagi
4. Menentukan Rouge atau Bleu

Prompt

**1. Persiapan Data:**

* Muat dataset utama yang memiliki kolom: no, nama, kalimat, tema\_postingan, ponpes, dan aspek\_manual.
* Fokuskan analisis pada kolom kalimat dan tema\_postingan.
* Muat dua file lexicon:
  + positive.csv: berisi kata positif dan bobot (word, weight)
  + negative.csv: berisi kata negatif dan bobot (word, weight)

**2. Visualisasi Awal:**

* Tampilkan **Wordcloud** dari kolom kalimat dan tema\_postingan **sebelum praproses**.

**3. Praproses Teks:**

Lakukan tahapan berikut untuk setiap teks:

1. **Cleaning** – hapus tanda baca, angka, karakter khusus.
2. **Deteksi Emoji** – ubah emoji menjadi token deskriptif (misal 😊 → emoji\_senyum).
3. **Case Folding** – ubah ke huruf kecil semua.
4. **Normalisasi** – ubah kata tidak baku menjadi baku.
5. **Tokenisasi** – pecah teks menjadi kata.
6. **Stopword Removal** – hilangkan kata umum.
7. **Stemming** – ubah ke bentuk dasar.

**4. Penggabungan Fitur Teks:**

* Gabungkan hasil praproses dari kalimat dan tema\_postingan menjadi satu fitur teks per entri.

**5. Analisis Frekuensi Kata:**

* Tampilkan daftar dan grafik frekuensi kata tertinggi dari teks yang telah diproses.

**6. Analisis Sentimen (Lexicon-Based):**

* Gunakan positive.csv dan negative.csv untuk memberi skor pada kata dalam teks:
  + Hitung total skor positif dan negatif.
  + Klasifikasikan sebagai:
    - **Positif** (skor > 0)
    - **Negatif** (skor < 0)
    - **Netral** (skor = 0)
* Simpan hasil ke kolom sentimen.

**7. Statistik & Visualisasi Sentimen:**

* Tampilkan jumlah masing-masing sentimen.
* Tampilkan **Wordcloud** untuk sentimen positif, netral, dan negatif.

**8. Topik Modeling (LDA):**

* Konversi teks ke representasi numerik (TF-IDF atau BoW).
* Terapkan **LDA** untuk mengelompokkan teks ke dalam **10 topik**.
* Pemetaan topik ke aspek:
  + Kualitas Guru
  + Sosial
  + Motivasi/Spiritual
  + Kegiatan Pondok
  + Akademik
  + Umum
  + Fasilitas
  + Lingkungan
  + Pembinaan Karakter
  + **Prestasi**
* Tetapkan topik dominan sebagai nilai di kolom aspek\_manual.

**✅ Penyimpanan Hasil Analisis Awal:**

* Setelah kolom aspek\_manual dan sentimen selesai terisi, simpan file hasil sebagai:  
  📄 **hasil\_sentimen\_pesantren.csv**

**9. Klasifikasi Sentimen Menggunakan RNN:**

**a. Split Dataset:**

* Pisahkan data menjadi:
  + Training: 80%
  + Testing: 20%
* Tampilkan grafik distribusi.

**b. Vektorisasi:**

* Ubah teks ke vektor menggunakan **TF-IDF**.

**c. Model RNN:**

* Bangun dan latih model RNN.
* Prediksi hasil testing.

**d. Evaluasi RNN:**

* Tampilkan **confusion matrix**.
* Hitung dan tampilkan **akurasi**.
* Tampilkan:
  + **Tabel nilai aktual vs prediksi**
  + **Grafik nilai aktual vs prediksi**

**10. Evaluasi Akhir & Penyempurnaan:**

* Validasi manual sebagian data.
* Lakukan penyempurnaan bila hasil kurang akurat (di preprocessing, topik, atau sentimen).