**NAMA : MUHAMMAD RIVALDI JEFRI**

**NIM : 221011041**

**TUGAS RETREIVAL INFORMATION**

**METODE EKTRAKTIF**

**Definisi**: Metode ekstraktif memilih kalimat atau bagian penting langsung dari teks asli tanpa mengubah kata-katanya. Kalimat dipilih berdasarkan skor kepentingan (misalnya, menggunakan algoritma seperti TF-IDF).

**Contoh Algoritma**: **TF-IDF** (Term Frequency-Inverse Document Frequency)

1. **TF (Term Frequency)**: Mengukur seberapa sering sebuah kata muncul dalam dokumen.
2. **IDF (Inverse Document Frequency)**: Mengurangi bobot kata yang muncul di banyak dokumen (kata umum seperti "dan" atau "yang").
3. **TF-IDF**: Menggabungkan TF dan IDF untuk memberikan skor kepentingan pada kata atau kalimat.

**Cara Kerja TF-IDF untuk Ringkasan Ekstraktif**:

1. Hitung skor TF-IDF untuk setiap kata dalam dokumen.
2. Berikan skor pada setiap kalimat berdasarkan jumlah skor TF-IDF kata-kata di dalamnya.
3. Pilih kalimat dengan skor tertinggi sebagai ringkasan.

Contoh Penerapan TF-IDF

**Teks Asli**: "Indonesia memiliki hutan hujan tropis yang kaya akan keanekaragaman hayati. Hutan ini menyediakan habitat bagi berbagai spesies flora dan fauna. Namun, deforestasi menjadi ancaman serius. Banyak pohon ditebang untuk lahan pertanian dan industri."

**Proses TF-IDF**:

1. Kata seperti "hutan", "keanekaragaman", "deforestasi" mendapat skor tinggi karena sering muncul dan spesifik.
2. Kata seperti "yang" atau "dan" mendapat skor rendah karena umum.
3. Kalimat "Indonesia memiliki hutan hujan tropis yang kaya akan keanekaragaman hayati." dan "Namun, deforestasi menjadi ancaman serius." mendapat skor tertinggi.

**Ringkasan Ekstraktif**: "Indonesia memiliki hutan hujan tropis yang kaya akan keanekaragaman hayati. Namun, deforestasi menjadi ancaman serius."