**TUGAS-1 : Data Mining Text**

**Kerjakan soal-soa berikut ini, tulis tangan. Program dan hasil ‘run’ program di-print dan dilampirkan dalam tugas.**

**(1) Jelaskan istilah berikut ini : NLP, NLTK, TF-IDF, LDA, LSA, Speech Acts.**

**(2) Jelaskan langkah-langkah dari “preprocessing text”, apa yang dihasilkan oleh setiap langkah pada proses tersebut.**

**(3) Siapkan sebuah dokumen text, tuliskan sumber-nya, sebaiknya dari salah satu Corpus dengan jumlah kata maksimum 100 kata (setiap mahasiswa harus menyiapkan dokumen text yang berbeda, tidak boleh sama, kalau ada yang sama maka nilainya akan dibagi bersama. Lakukan preprocessing terhadap dokumen tersebut secara lengkap. Print program dan hasil proses-nya untuk dilampirkan.**

**(4) Tuliskan perbedaan antara “Topic Mining” dan “Probabilistic Topic Models”.**

**(5). Disediakan document berupa judul beberapa buku berikut ini:**

**documents = [ “Problem Solving using Python”,**

**“Introduction to Science through Python Programs”,**

**“Algorithm and Data Structure in Python”**

**“ Python Data Science Explained Easily” ]**

**(a) Lakukan “Topic Mining” terhadap documents dengan memilih salah satu teknik topic mining.**

**(b) Gunakan salah satu teknik “Probabilistic Topic Models” untuk mengolah documents tersebut.**

**(c) Hitung nilai TF-IDF dari kata “Python” dalam documnets tersebut.**

**KUMPULKAN TUGAS-1 INI PADA HARI KETIKA MID DILAKSANAKAN! .**