Nama : Muhamad Faqih Azhar

NIM : 0110221092

Kelas : 5TI03 / Senin Malam

**Repository**

*<https://github.com/FaqihAzh/nlp-task/tree/main/week-4>*

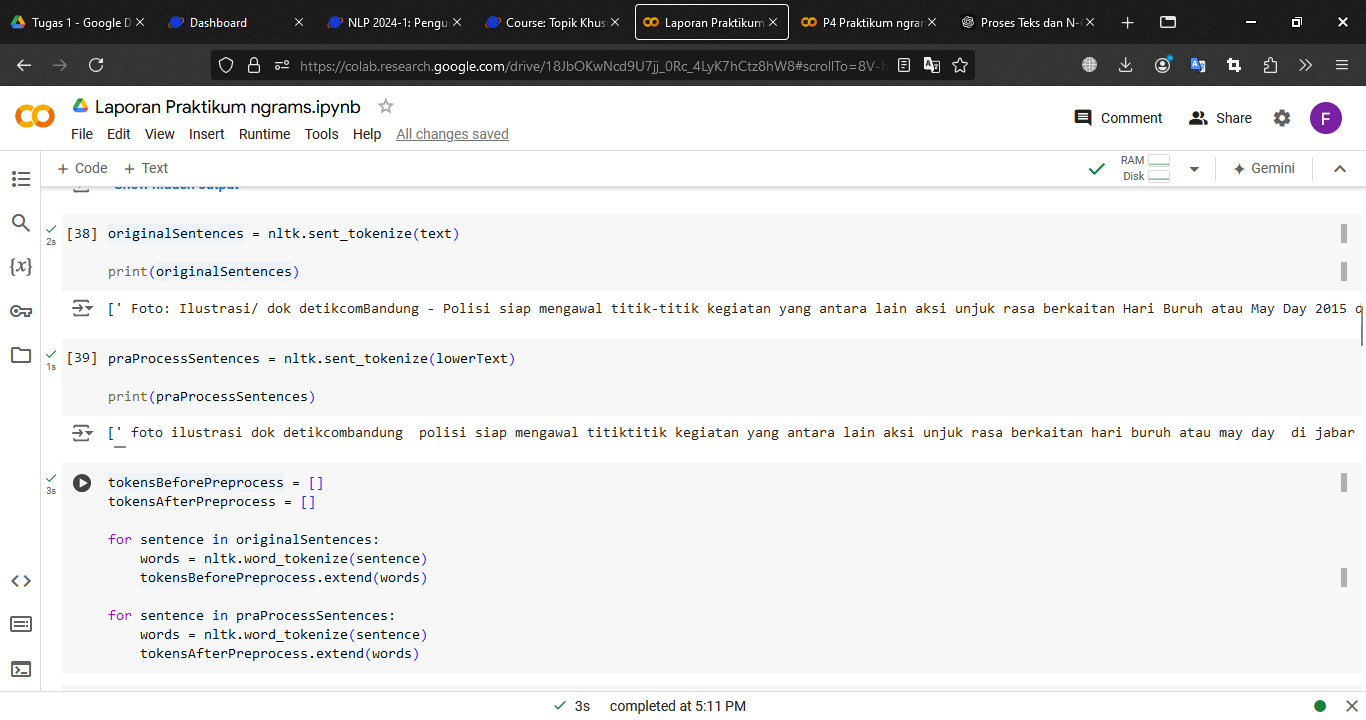
**Pertanyaan**

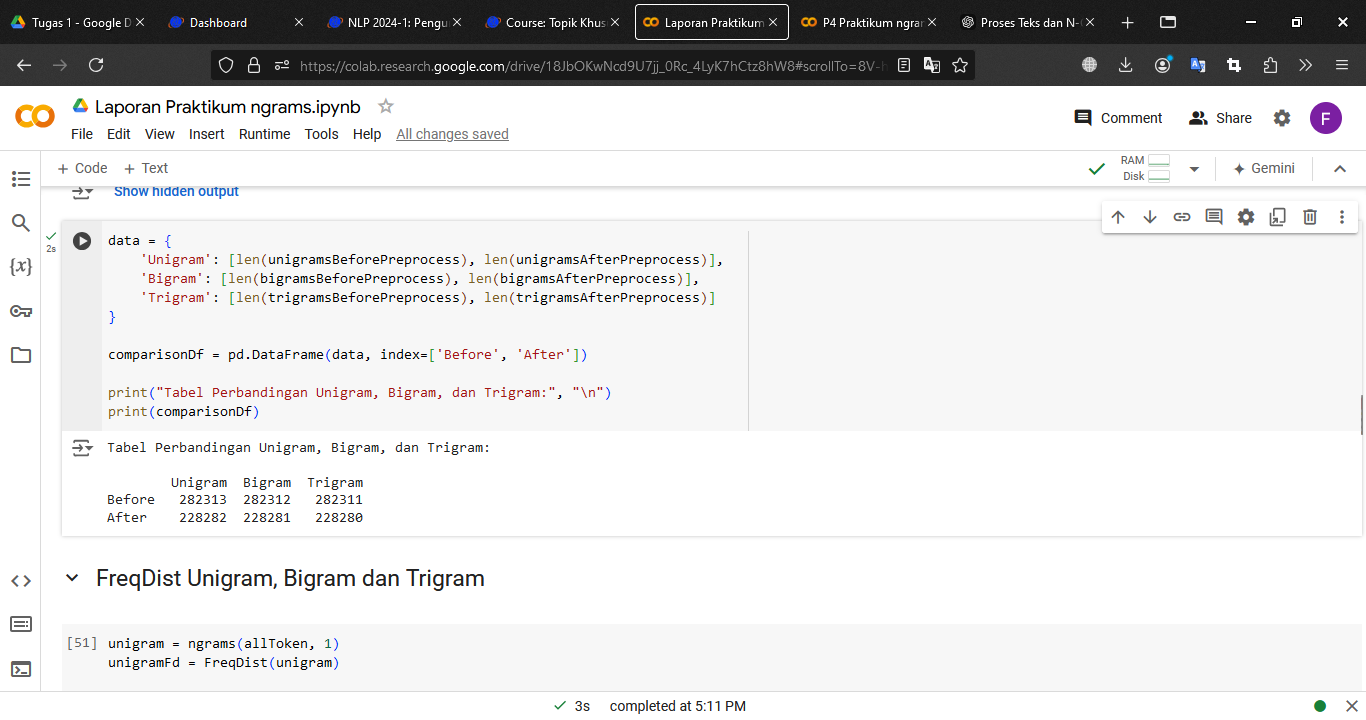
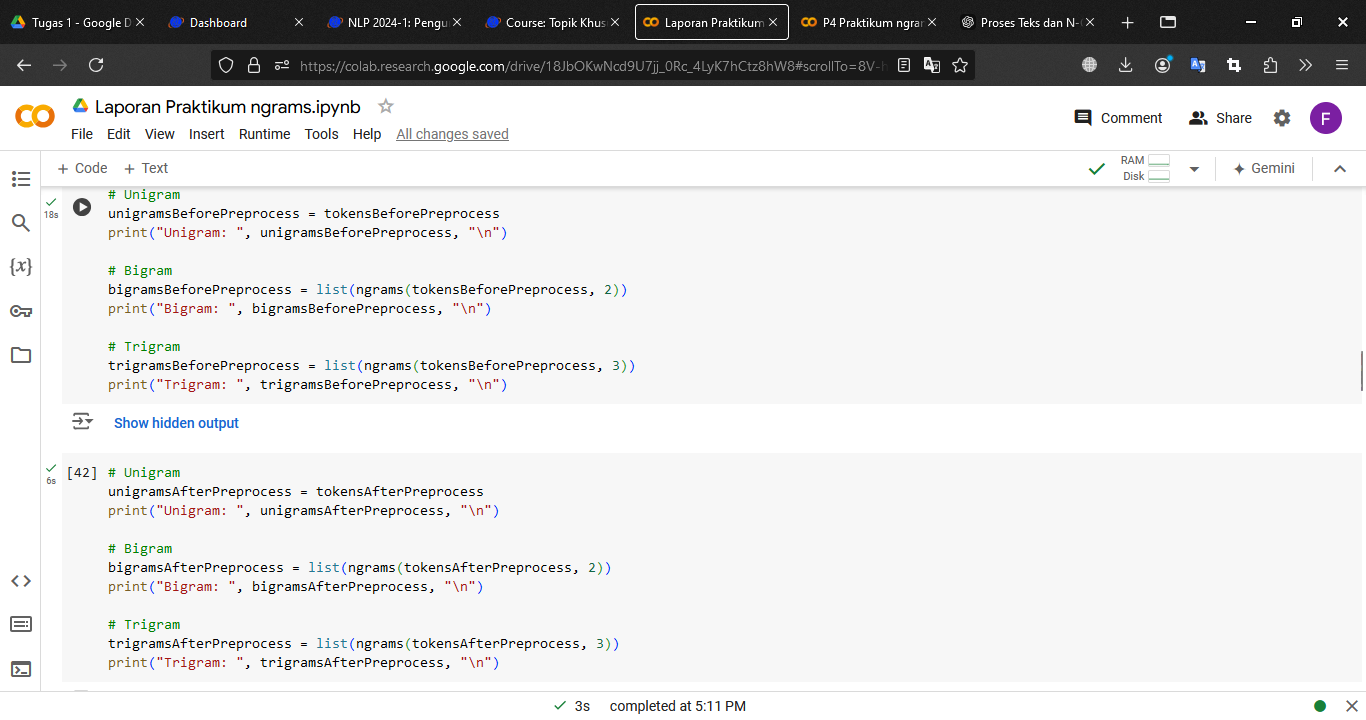
1. Lakukan pra-proses terhadap data teksnya terlebih dahulu
2. Hapus simbol, angka, dan token yang tidak umum,
3. Ubah semua teks menjadi lowercase

**Jawab:**

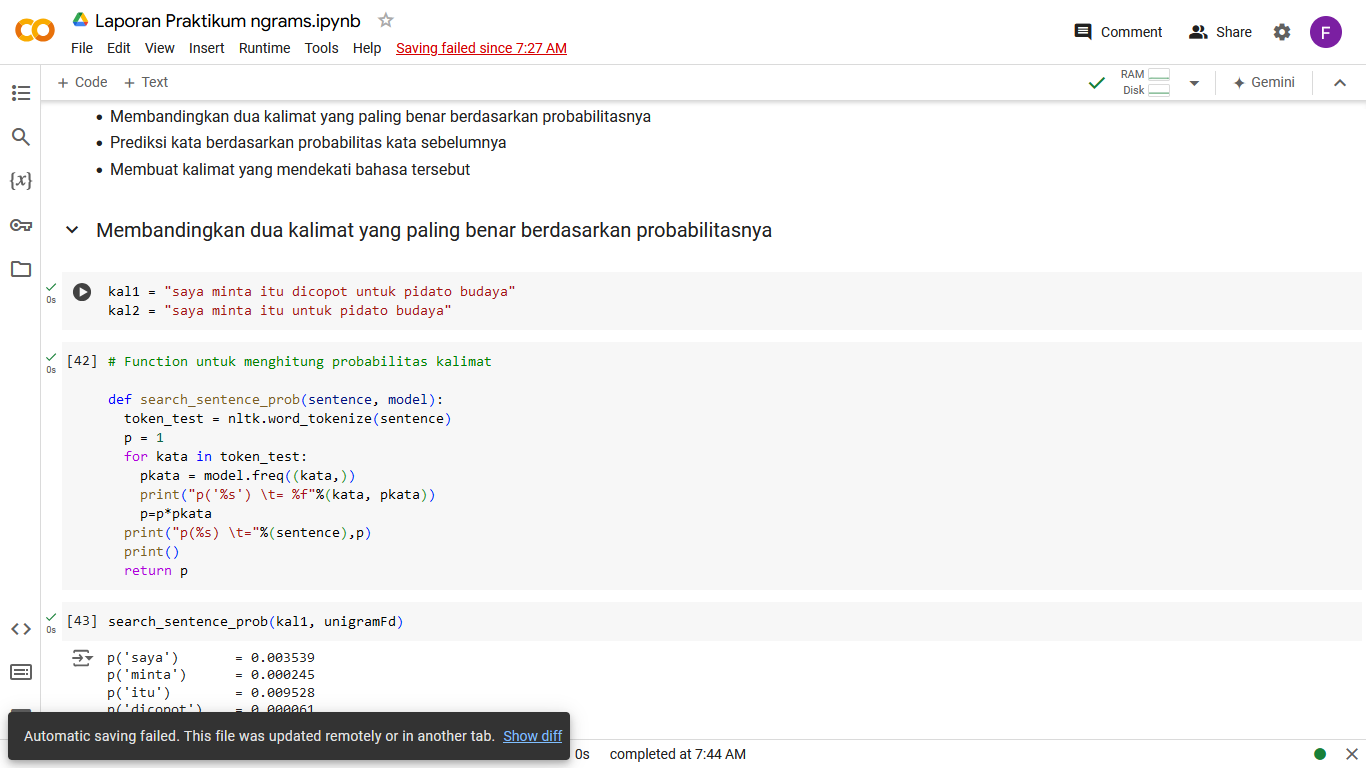


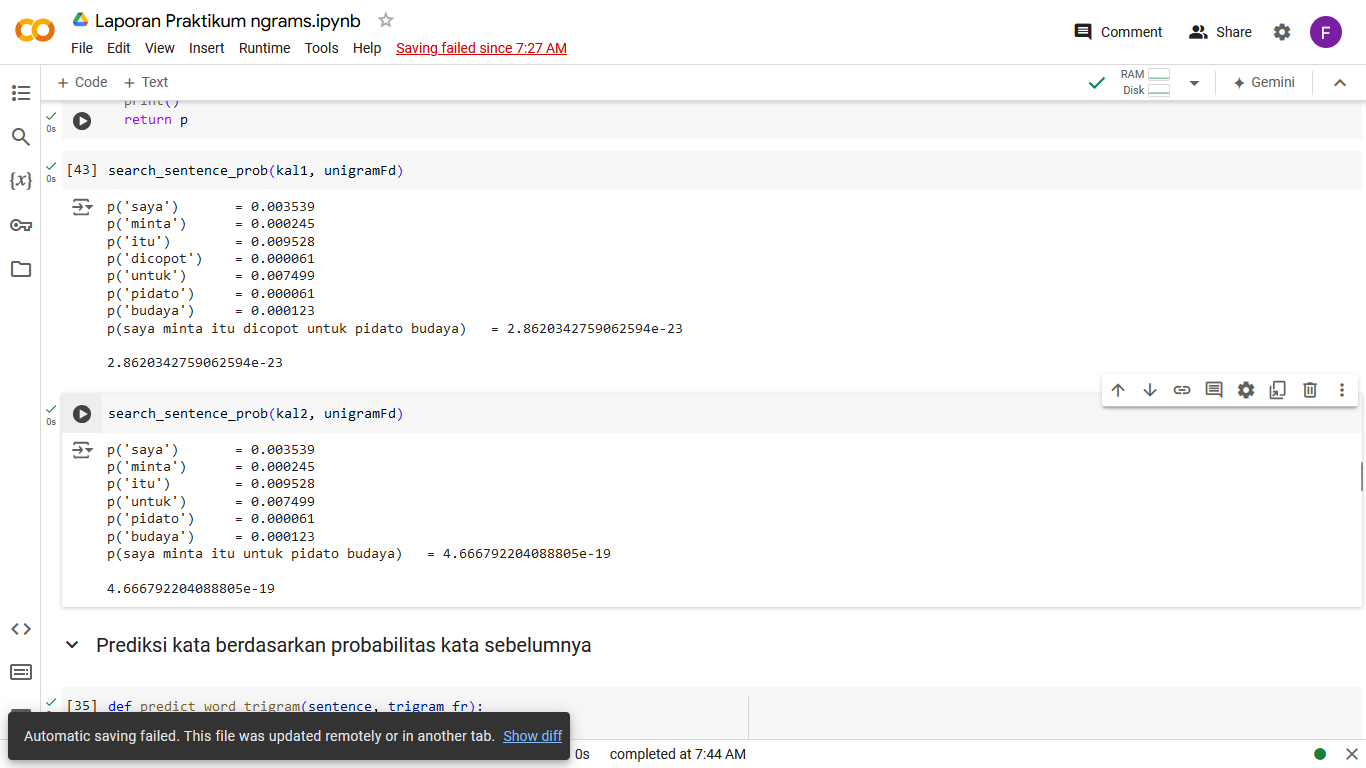
1. Hitung berapa banyak token unigram, bigram, dan trigramnya sekarang kemudian bandingkan dengan kondisi awal yang sebelum pra-proses!



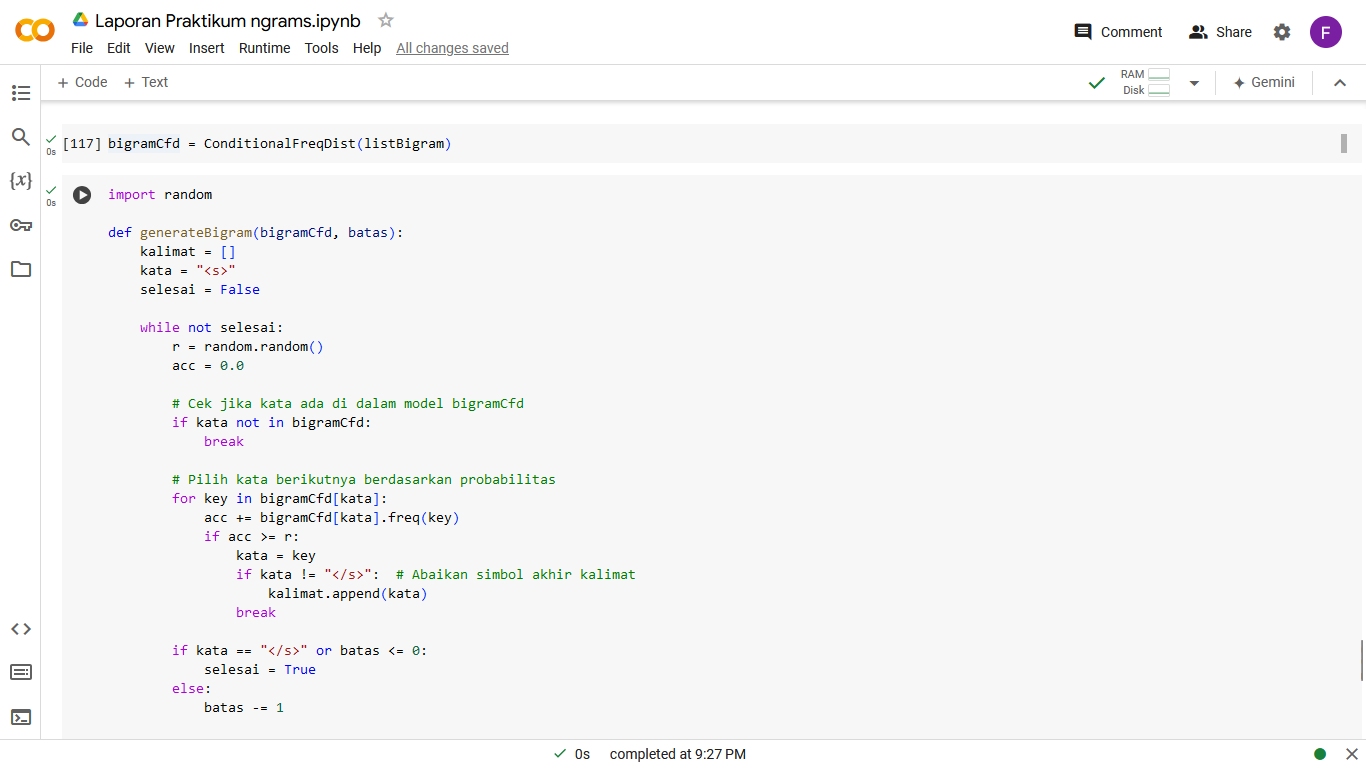


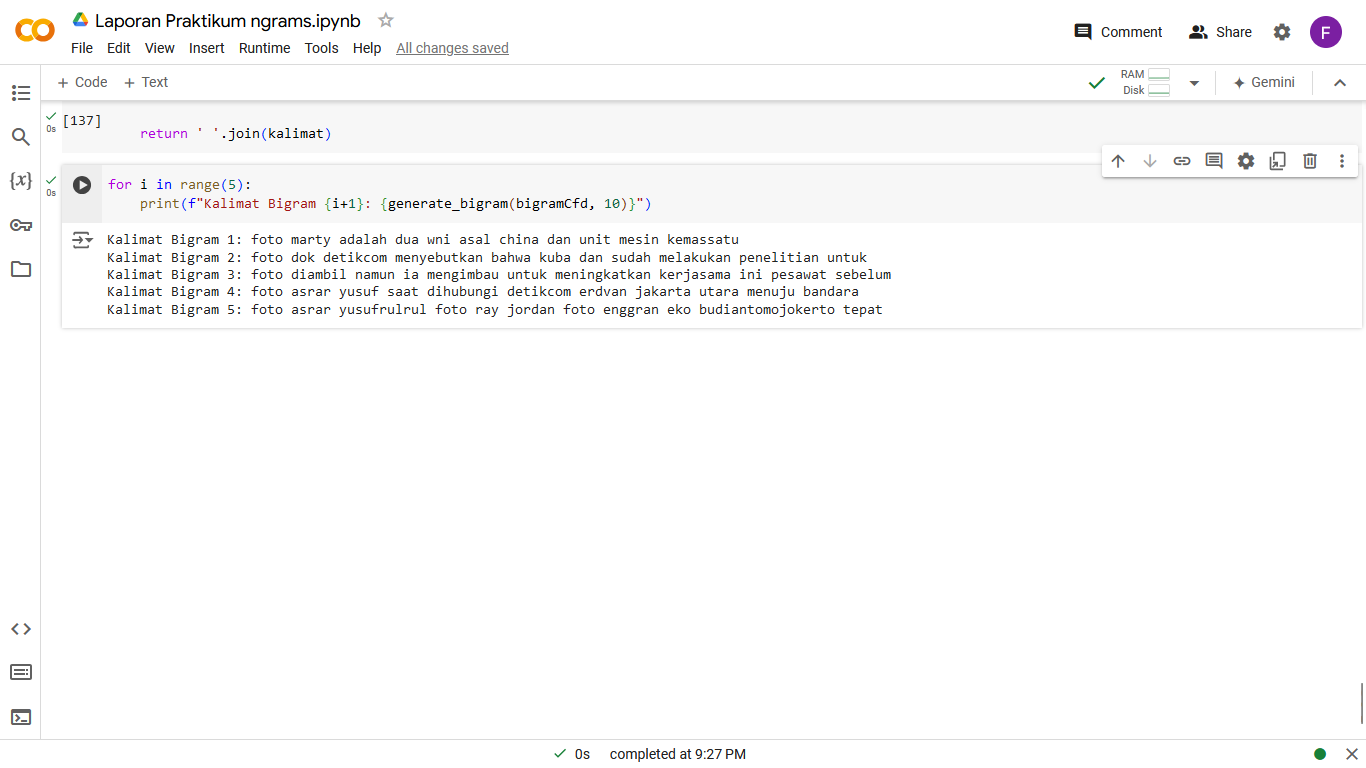
1. Hitung ulang peluang dari kedua kalimat yang diberikan (kal1 dan kal2), mana yang lebih mungkin untuk terjadi?





1. *Generate* 5 buah kalimat dari corpus berita yang diberikan! Diperbolehkan memanfaatkan bigram, trigram, atau ngram di atasnya.





1. Buat sebuah pantun 2 bait dari corpus pantun dan diperbolehkan memberikan syarat tambahan sesuai dengan masing-masing!

