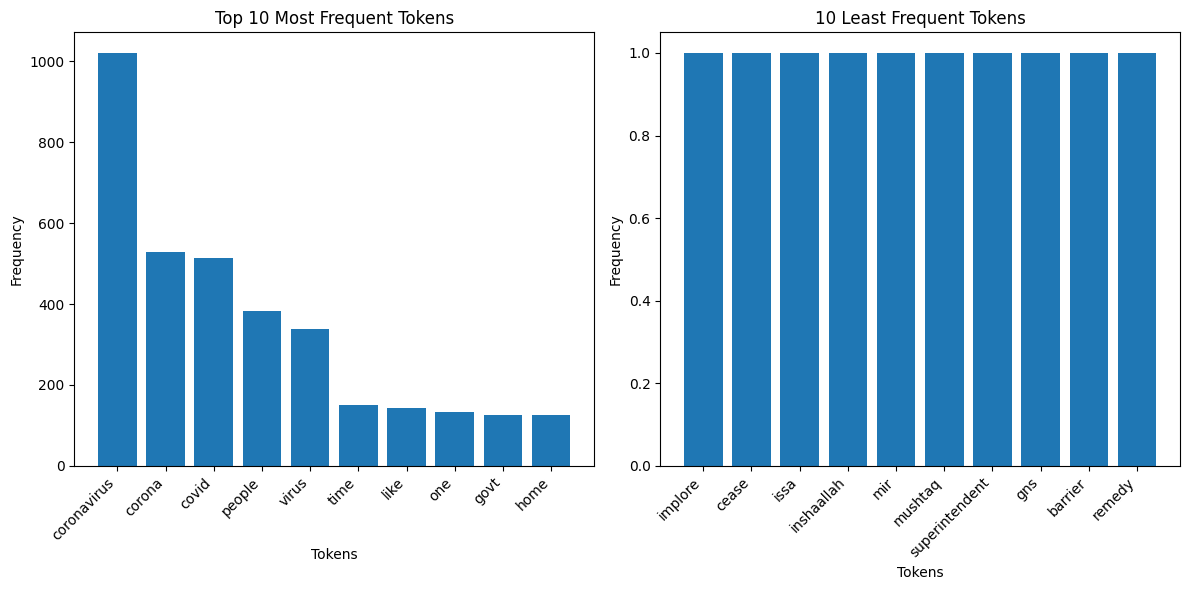
**گزارش تمرین 3 یادگیری ماشین**

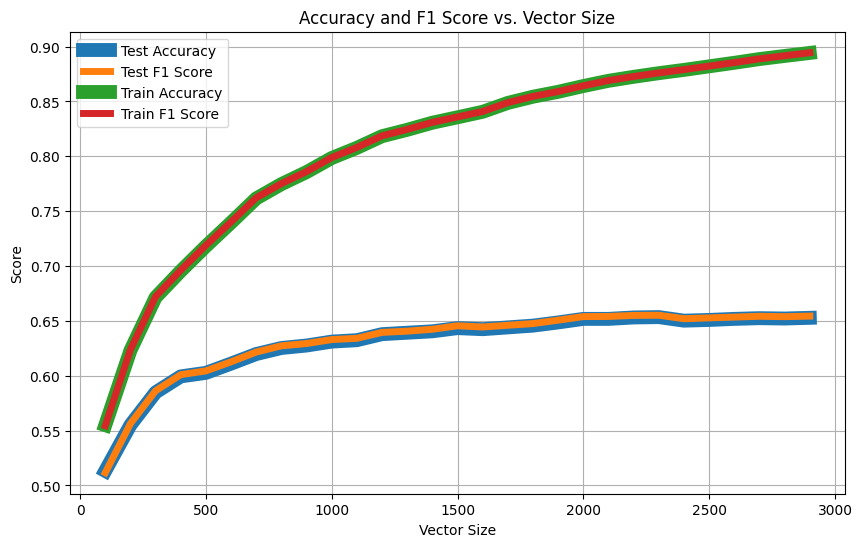
**مازندرانیان**

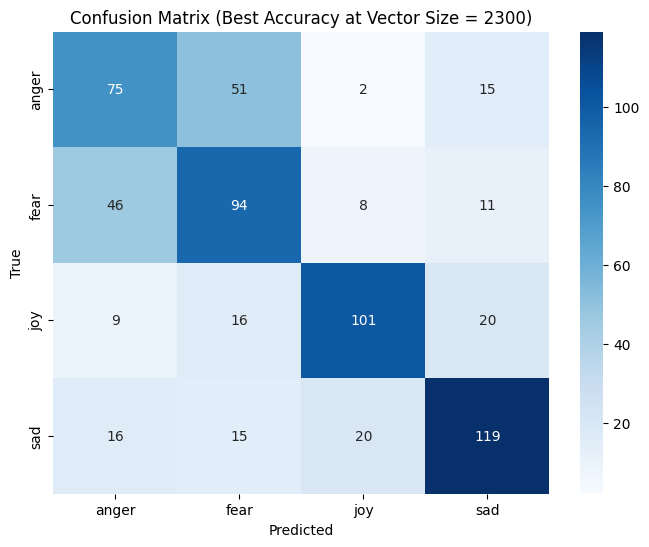
**830402066**

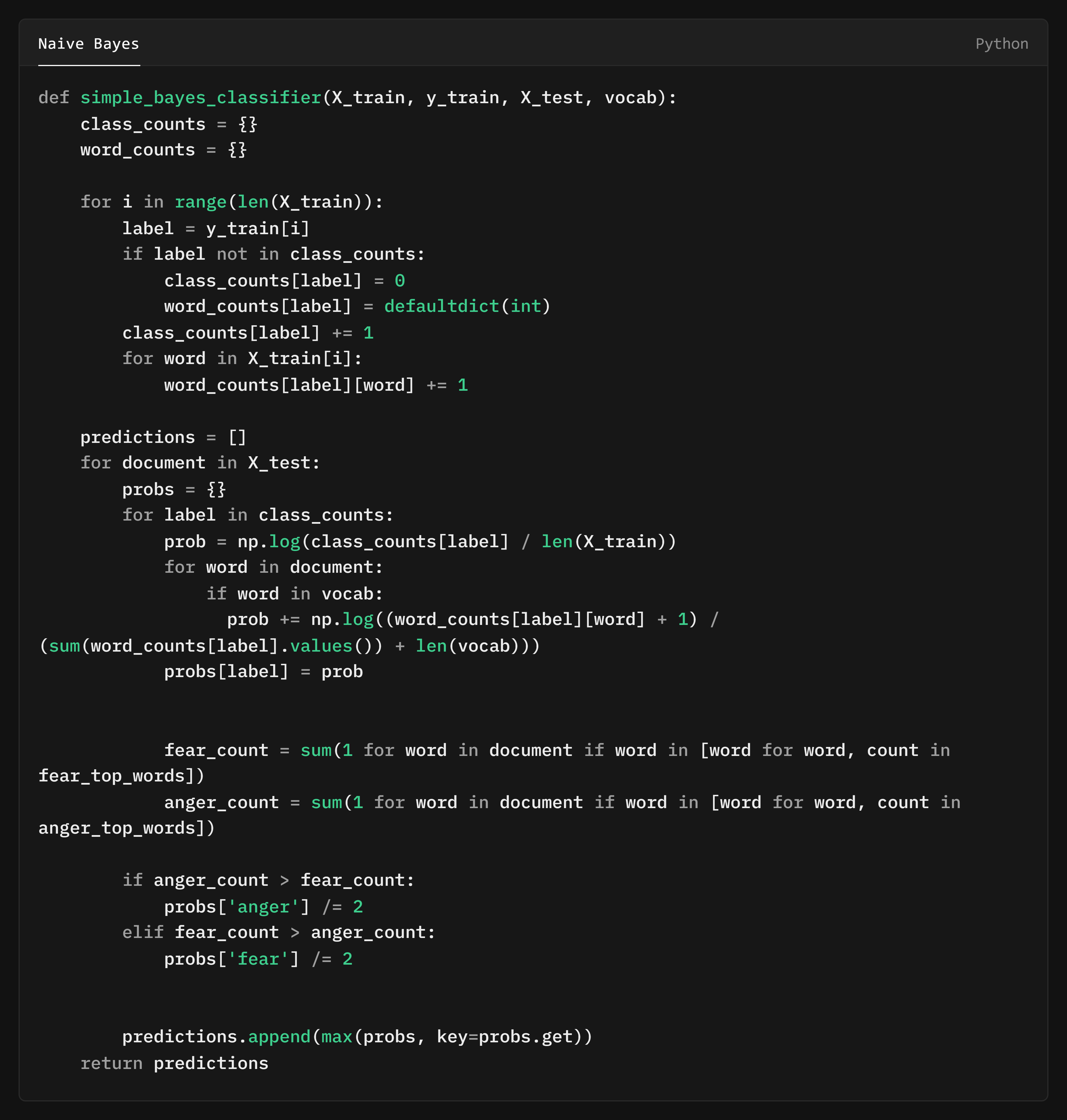
**سوال 1)**

ابندا طبق خواسته های سوال پیاده سازی انجام شد ولی در بهترین حالت و برای vector\_size=3000 دقتی برابر با 65 درصد استخراج شد. همچنین با کتابخانه ی sklearn و روش TFIDF که معمولا روش بهتری نسبت به Bag Of Words است هم به دقت بیشتری دست پیدا نکردیم. عمده مشکل این دقت پایین برای کلاس های anger و fear بود که مدل پیش بینی های این کلاس ها رو اکثرا جابجا انجام میداد. به عنوان روش جایگزین کلماتی را که بین 4 کلاس برای هر کدام متمایز هستند را در متن نگه میداریم و مابقی کلمات را حذف میکنیم. همچنین با استفاده از کتابخانه ی nltk کلمات stop words هم حذف شدند. این کار منجر به این شد که دفت مدل 100 درصد شود که البته نشان دهنده ی overfit بودن روی این دیتاست است.



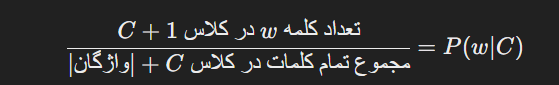




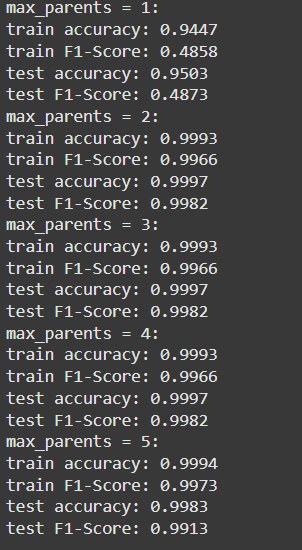


این تابع یک مدل ساده‌ی بیزین برای دسته‌بندی متون پیاده‌سازی کرده است. هدف آن پیش‌بینی برچسب مناسب برای هر سند جدید بر اساس داده‌های آموزشی است. ابتدا داده‌های آموزشی را بررسی کرده و تعداد اسناد مربوط به هر برچسب (کلاس) و تعداد تکرار کلمات در هر برچسب را می‌شمارد. این اطلاعات برای محاسبه احتمال وقوع هر کلاس و احتمال شرطی کلمات در آن کلاس استفاده می‌شود.

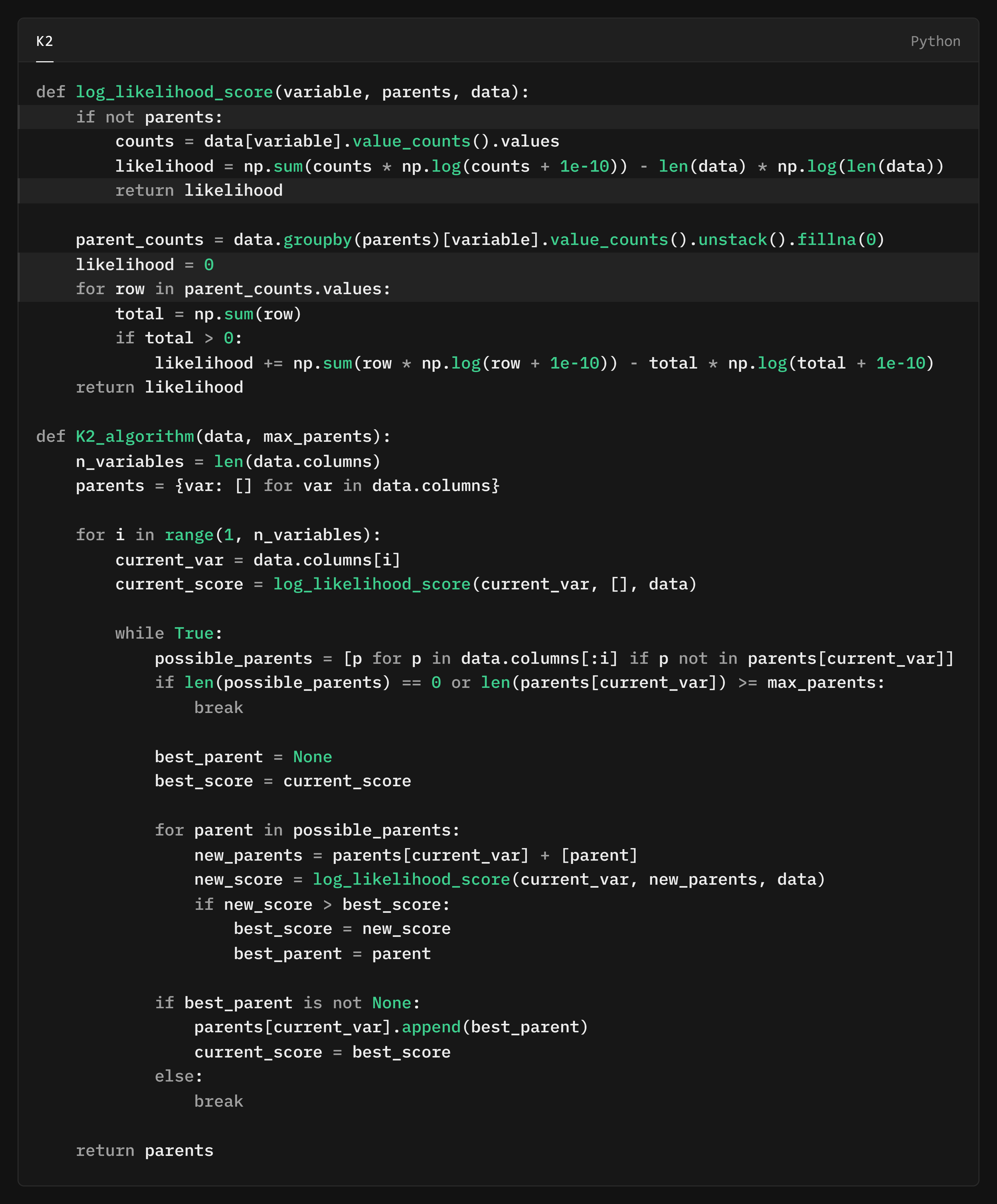
سپس برای اسناد جدید، احتمال تعلق به هر کلاس با استفاده از احتمالات محاسبه‌شده قبلی تخمین زده می‌شود. در این فرآیند، اگر کلمات مربوط به احساسات خاصی (مثل ترس یا خشم) در متن وجود داشته باشند، احتمال کلاس مرتبط با آن احساس کاهش می‌یابد. در نهایت، کلاسی که بیشترین احتمال را دارد به عنوان پیش‌بینی برچسب متن انتخاب می‌شود.



**سوال 2)**

****

****



در ابتدا، برای هر متغیر یک امتیاز اولیه محاسبه می‌شود که نشان می‌دهد آن متغیر بدون والدین چه قدر خوب مدل‌سازی می‌شود. سپس الگوریتم شروع به جست‌وجوی بهترین والدین برای هر متغیر می‌کند. والدین یک متغیر، متغیرهایی هستند که تأثیر مستقیم روی آن دارند. برای پیدا کردن والدین، الگوریتم والدینی را اضافه می‌کند که بیشترین بهبود را در امتیاز متغیر ایجاد کنند. این کار ادامه پیدا می‌کند تا وقتی که یا هیچ بهبودی در امتیاز حاصل نشود یا تعداد والدین از مقدار مجاز بیشتر نشود.

در نهایت، خروجی این الگوریتم یک ساختار شبکه است که برای هر متغیر، لیستی از والدینش را نشان می‌دهد. این ساختار می‌تواند در مسائل مختلفی مثل پیش‌بینی و تحلیل روابط بین متغیرها استفاده شود.

