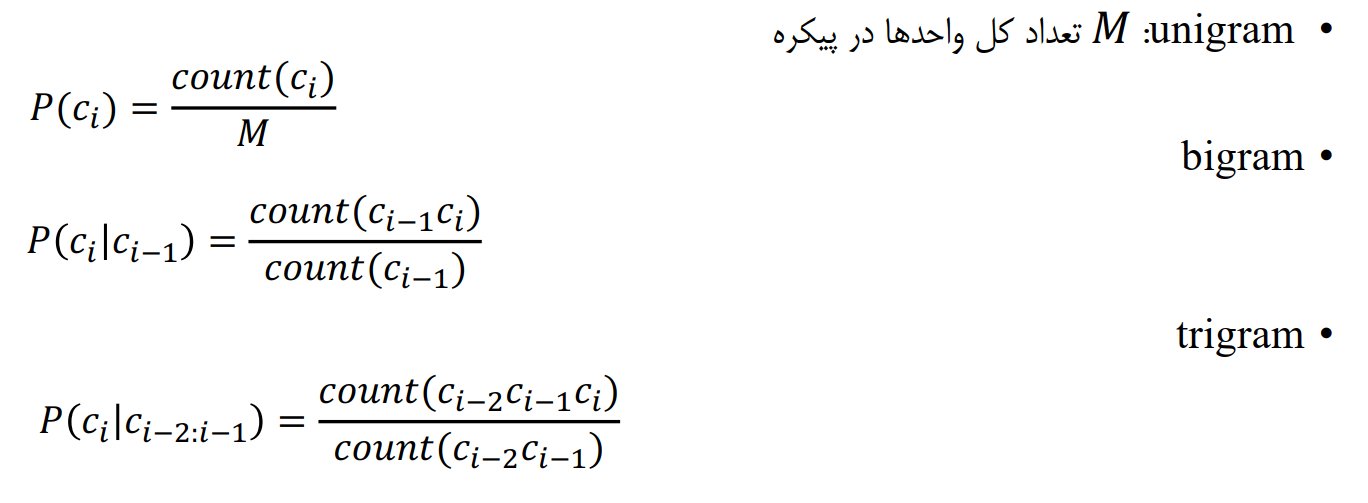
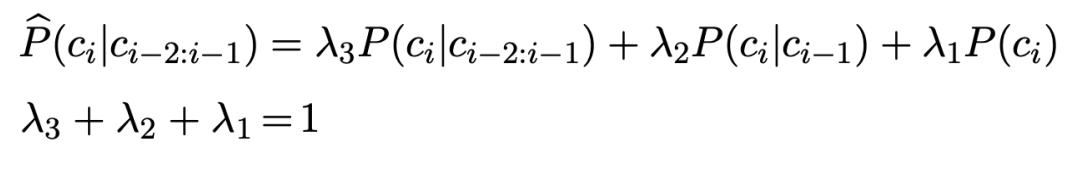
**توضیحات:**

ابتدا از فایل 'Train\_data.txt' جملات داده‌شده را خوانده، سپس با استفاده از تابع word\_tokenize، کلمات هر جمله را استخراج می‌کنیم. تمام کلمات را در یک set نگه‌داری می‌کنیم. تعداد هر کلمه به طور مجزا، زوج کلمه و ۳ تایی بررسی‌ شده و ذخیره می‌شوند.

برای هر چندتایی از کلمات یک تابع محاسبه احتمال با فرمول های زیر در نظر گرفته‌شده است:



تابع guess\_next\_word کلمه‌ی جا افتاده را بر اساس کلمات‌ ذخیره‌شده حدس می‌زند.



0.0000130.009870.990117

اعداد تصادفی انتخاب شده‌اند و برای واقعی‌تر بودن نتایج ضریب Trigram از بقیه بیشتر است.

جملات ورودی از فایل Train\_data خوانده‌شده و تابع guess\_next\_word روی $ اعمال می‌شود؛ سپس نتایج نمایش داده شده، در فایل Test\_data\_result ذخیره‌شده و با فایل labels مقایسه می‌شوند.

دقت مدل امتحان‌شده: 0