**ШИНЖЛЭХ УХААН ТЕХНОЛОГИЙН ИХ СУРГУУЛЬ**

**Мэдээлэл холбооны технологийн сургууль**



**БИЕ ДААЛТЫН АЖЛЫН**

**ТАЙЛАН**

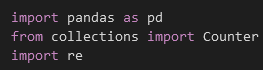
# **Эх Хэлний Боловсруулалт (F.CSC336) 2024-2025 оны хичээлийн жилийн намар**

**Бие даалтын ажлын нэр:** **Корпус үүсгэн хэлний загвар байгуулах**

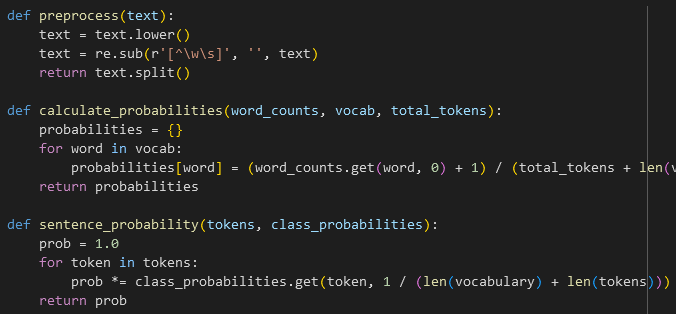
# **Хичээл заасан багш:** Э.Батцэцэг

# **Бие даалтын ажил гүйцэтгэсэн:** С.Тэмүүжин B221960002

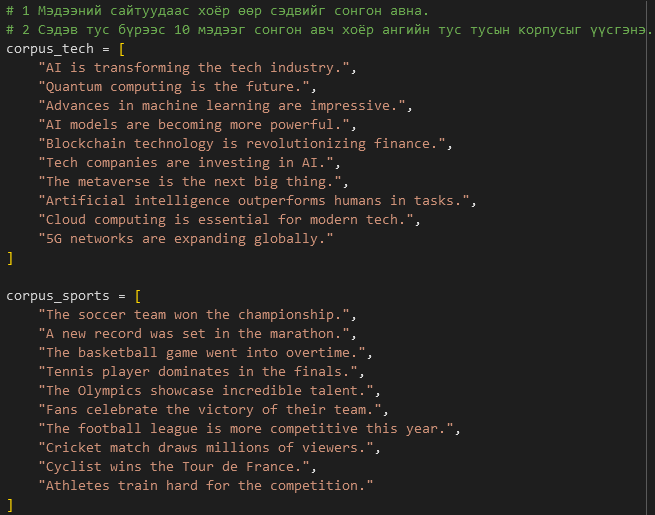
**Улаанбаатар хот. 2024**



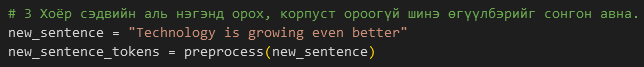
Эдгээр функцууд нь текстийг цэвэрлэж, тэмдэглэгээ хийх замаар урьдчилан боловсруулж, өгөгдсөн ангийн үгийн магадлалыг Лаплас тэгшитгэгчээр (calculate\_probabilities) тооцоолж, тухайн ангид хамаарах өгүүлбэрийн хуримтлагдах магадлалыг тус тусад нь үгийн магадлалыг (sentence\_probability) үржүүлж тооцдог. Тэд хамтдаа Байесын ангиллын гол цөмийг бүрдүүлдэг.



Техник, Спорт гэсэн хоёр өөр сэдвийг сонгодог. Сэдэв бүрийн хувьд бид 10 өгүүлбэр (мэдээ эсвэл синтетик) сонгоно. Эдгээр өгүүлбэрүүд нь анги бүрийн корпусыг бүрдүүлдэг. Эдгээр корпус дахь үгс нь үгийн магадлалыг тодорхойлно.



Бид сэдвүүдийн аль нэгэнд багтах боловч үндсэн хэсэгт ороогүй шинэ өгүүлбэрийг сонгодог.



Өгүүлбэр бүрийг жижиг үсгээр хувааж, цэг таслалыг арилга.

Жишээ:

"Technology is growing even better." ["technology", "is", "growing", "even", "better”] болно.

Анги тус бүрийн корпуст үг бүр хэр олон удаа гарч байгааг тоол.

Жишээ:

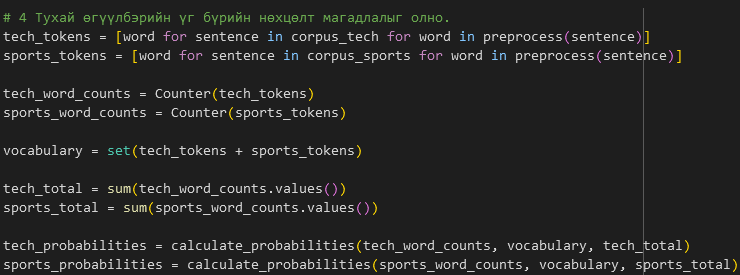
Техникийн корпус дээр "technology" нь 1 удаа гарч ирж болно.

Тайлбар толь үүсгэх:

Хоёр ангийн бүх өвөрмөц үгсийг үгсийн сан болгон нэгтгэ. Энэ нь ангиудыг харьцуулах үед үгийн магадлалыг тогтвортой байлгах боломжийг олгодог.

Лаплас гөлгөр болгох:

Тоолох үг бүрт 1-г нэмнэ



Анги тус бүрийн магадлалыг ашиглан бүх өгүүлбэрийн магадлалыг тооцоолно.

Үгний магадлал:

Шинэ өгүүлбэр дэх үг бүрийн магадлалыг Техникийн болон Спортын ангиудад олоорой.

Жишээ:

Техник дэх "technology"-ын магадлал = 0.013514

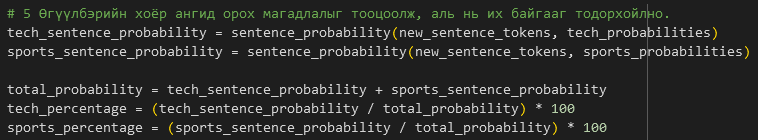
Спорт дахь "technology"-ын магадлал = 0.006536

Өгүүлбэрийн магадлал:

Анги тус бүрийн бүх үгсийн магадлалыг үржүүлнэ (бие даасан гэж үзвэл).

100% хүртэл хэвийн болгох:

Эдгээр магадлалыг тус бүрийг хоёр магадлалын нийлбэрт хуваах замаар хувь болгон хөрвүүл.



Илүү өндөр магадлалтай анги бол таамагласан сэдэв юм.

Жишээ:

Техникийн магадлал = 91.44%

Спортын магадлал = 8.56%

Урьдчилан таамаглах = Техник

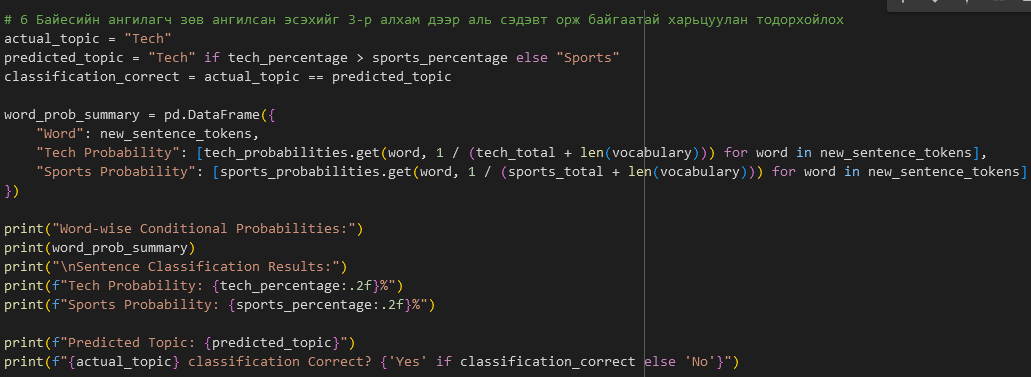
Зөв байдлыг шалгах:

Урьдчилан таамагласан сэдвийг бодит сэдэвтэй харьцуул.

Жишээ:

Бодит сэдэв = Техник

Хэрэв таамагласан болон бодит сэдвүүд таарч байвал ангилал зөв байна.



Зураг дээр Техник болон Спортын ангиудад магадлалыг тооцсон өгүүлбэрийг үгээр нь шинжлэн дүгнэсэн Байес ангиллын үр дүнг харуулж байна. Ангилагч нь өгүүлбэрийг 86.51% -ийн итгэлтэйгээр Техникийн ангилалд хамаарах гэж таамаглаж байгаа бөгөөд ангилал зөв байна.

