

Bayes

Kodutöö 7

Programmi eesmärk on teha kindlaks, kas antud teksti puhul on tegemist spämmiga või tavalise kirjaga. Selleks peab vaatama palju on sõnu ja mis sõnad on tekstis ning nende põhjal otsustada, mis tekstiga on tegu. Funktsioon **read_directory()** loeb näite faile erinevate tekstiga kaustast. Peamised funktsioonid töös on **calculate_ham_chance()** ja **calculate_spam_chance()**. Nende ülesanne on vastavalt arvutada, mis tõenäosusega on tekst spämm. Nendes vaadatakse dictionary, kui mitu korda esines teatud sõnad, mis on kujul sõna on key ja esinemis kordi on value. Lisaks vaadatakse palju on tähti selles tekstis ja jagatakse 6000. Pärast seda kutsutakse välja **calculate_probablity_of_word()**, millega kutsutakse välja koos eelnevalt arvutatud väärtustega. Selles funktsioonis unikaalsete sõnade arvu põhjal ja pikkuse põhjal antakse tagasi tulemus funktsioonis **solve()**, mis tagastab True, kui on spämm ja False kui ei ole.

```
Ham chance  2.663051774583386
Spam chance  2.5143977096822527
Not spam

Ham chance  3.0334881058389507
Spam chance  3.7396261588907933
```