

**Hasan Suca Kayman**

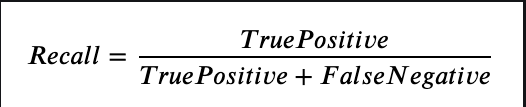
**2017280030**

**BIL3102-Visa-S2**

# Aşağıda verilmiş olan tabloda, 3 farklı sınıf (A, B, C) için her bir kaydın gerçek ait olduğu sınıf ve algoritmanın tahmin ettiği sınıf bilgileri bulunmaktadır. Bu tablodan giderek aşağıdaki şıkları yanıtlayınız.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Algoritmanın Tahmin Ettiği Sınıf** | | |
|  | | **A** | **B** | **C** |
| **Gerçek Sınıf** | **A** | 8 | 4 | 6 |
| **B** | 3 | 9 | 1 |
| **C** | 0 | 2 | 5 |

## **A** sınıfının **Recall** (Duyarlılık) değeri nedir**?**

A icin Recall:

Sari: True Positives

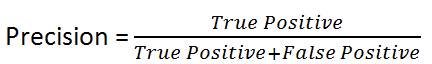
Yesil: False Positives

Mavi: False Negatives

Kirmizi: True Negatives

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Algoritmanın Tahmin Ettiği Sınıf** | | |
|  | | **A** | **B** | **C** |
| **Gerçek Sınıf** | **A** | 8 | 4 | 6 |
| **B** | 3 | 9 | 1 |
| **C** | 0 | 2 | 5 |

## **B** sınıfının **Precision** (Kesinlik) değeri nedir**?**

B icin Precision:

Sari: True Positives

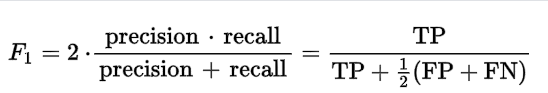
Yesil: False Positives

Mavi: False Negatives

Kirmizi: True Negatives

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Algoritmanın Tahmin Ettiği Sınıf** | | |
|  | | **B** | **A** | **C** |
| **Gerçek Sınıf** | **B** | 9 | 3 | 1 |
| **A** | 4 | 8 | 6 |
| **C** | 2 | 0 | 5 |

## **C** sınıfının **F-Score** (F1 Score) değeri nedir?

C icin F-score:

Sari: True Positives

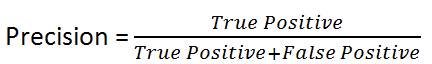
Yesil: False Positives

Mavi: False Negatives

Kirmizi: True Negatives

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Algoritmanın Tahmin Ettiği Sınıf** | | |
|  | | **C** | **B** | **A** |
| **Gerçek Sınıf** | **C** | 5 | 2 | 0 |
| **B** | 1 | 9 | 3 |
| **A** | 6 | 4 | 8 |

## **1.4)** Üç sınıfın **Precision** için **micro average** (mikro ortalama) değeri nedir?

A icin Precision:

B icin Precision:

C icin Precision:

[**Refences**](https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826)

# **2) N-gram ile dil modelleme analizi** yapıyorsunuz. Aşağıdaki tablolarda, bazı yazışmalardan elde edilen sözcüklerin unigram ve bigram dağılım tabloları (adet / count değerleri şeklinde) verilmektedir. Yazışmalara ait **V** yani Vocabulary (sözcük dağarcığı) değeri **= 165**.

Unigram tablosu

Bigram tablosu



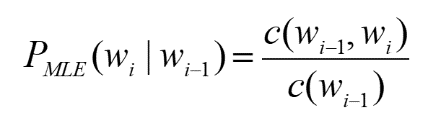
**Maximum Likelihood Estimate** yöntemi ile aşağıdaki **cümlenin bigram tahmini olasılığını** bulunuz.

çok büyük zarar yok

## **Normalizasyon Bigram:**

## **Cümleyi 2-gram(Bigram) ve Bayes Olasılık Halinde formüle aktarmalıyız:**

**P(<s>çok büyük zarar yok</s>)** = **P(**çok**│**<s>**)**×**P(**büyük**│**çok**)×P(**zarar**│**büyük**)×P(**yok**│**zarar**) ×P(**</s>**│**yok**)**

**P(** çok **│**<s>**)**= = = **.719**

**P(** büyük **│**çok**)= .063**

**P(** zarar **│**büyük**)=.066**

**P(** yok **│**zarar**)=.05**

**P(** </s> **│**yok**)=** = = **.75**

**P(<s> çok büyük zarar yok </s>)** = .719×.75×.063×.066×.05 =**0.00011**

**P(** çok **│**<s>**)**= hocam bunu kendi kesfettigim ve dogru dusundugum mantikla tahmin ettim. Bigramda “çok” kolonunun değerlerinin hepsini toplarsak mantıken “çok” kelimesinin öncesinde o kadar kelime var demektir. Bu elde ettiğimiz veriyi Unigram değerinden çıkarırsak öncesinde kelime olmayan(cümlenin başı) “çok” frekansını elde etmiş oluruz.

**P(** </s> **│**yok**)=** hocam bunu da kendi kesfettigim ve dogru dusundugum mantikla tahmin ettim. Bigramda “yok” sütununun değerlerinin hepsini toplarsak mantıken “çok” kelimesinin sonrasında o kadar kelime var demektir. Bu elde ettiğimiz veriyi Unigram değerinden çıkarırsak sonrasında kelime olmayan(cümlenin sonu) “çok” frekansını elde etmiş oluruz.

0 gibi bir problem çıkmadığı için Laplace\_smoothed kullanmadım hocam