**KARAKTER N-GRAM TABANLI DOKÜMAN SINIFLANDIRMA**

Özdenur UÇAR ve Alperen ÇAKIROĞLU

Bilgisayar Mühendisliği Bölümü

Kocaeli Üniversitesi

[ozdenuru@gmail](mailto:ozdenuru@gmail), alperenckroglu204@gmail.com

# Problem Tanımı

*Hangi konu hakkında yazıldığı bilinmeyen haberleri bilgisayarların gücü ve yazılımların zekası ve matematiği kullanarak konularını bulabiliriz. Yazılan program tam anlamıyla konusunu bulmasa da kategorisini bulabilmektedir. Bunun için 2 ve 3 gramlarına bölünmüş haberler üzerinde naive bayes algoritmalarını çalıştırır. Programın haber kategorilerini tahmin etmeden önce eğitilmesi gerekmektedir. Bu işlem 1150 tane haberin yüzde yetmiş beşi ile yapılır. Geri kalan yüzde yirmi beşlik kısım programı test etmek için kullanılmıştır. Projede kullanılan algoritma naive bayes algoritmasıdır. Tahminler naive bayes algoritmasına göre yapılır.*

# Yapılan Araştırmalar

# Tasarım

## Akış Şeması

## Yazılım Mimarisi

## Veri Tabanı Diyagramı

# 4. Genel Yapı

# 5.Referanslar