

RANDOM FOREST ALGORİTMASI

Ezgi ÇELİK

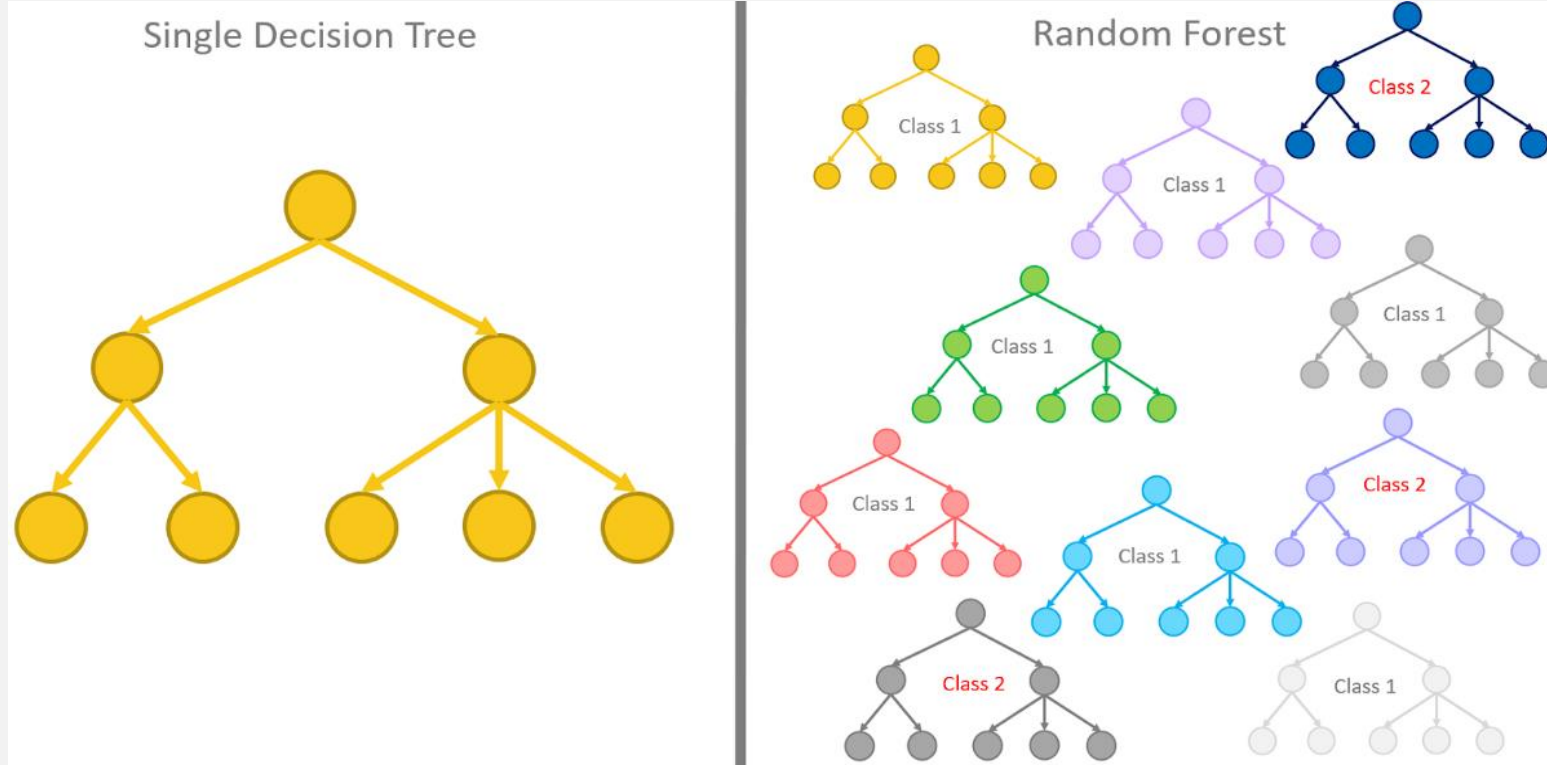
Ceyda AKTEPE

Özge TOKMAK

Ekrem Serdar ÖZTÜRK

Melik Atakan KEMER

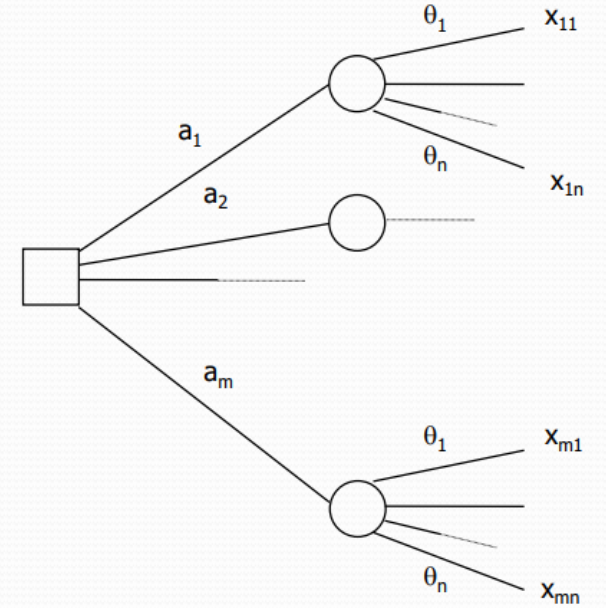
- Makine öğrenmesi modellerinden biri.
- Sınıflandırma işlemi esnasında birden fazla karar ağacı üreterek sınıflandırma değerini yükseltmeyi hedefleyen bir algoritmadır. Bireysel olarak oluşturulan karar ağaçları bir araya gelerek karar ormanı oluşturur. Buradaki karar ağaçları bağlı olduğu veri setinden rastgele seçilmiş birer alt kümedir.



KARAR AĞAÇLARI

- Bir karar ağacı aşağıdakilerden oluşan bir şekildir:
- Karar noktaları (kareler)
- Şans noktaları (daireler)
- Karar dalları (seçenekler)
- Şans dalları (olaylar)
- Son noktalar (getiriler veya faydalar)

SEÇENEKLER	OLAYLAR			
	θ_1	θ_2	...	θ_n
a_1	x_{11}	x_{12}	...	x_{1n}
a_2	x_{21}	x_{22}	...	x_{2n}
.
a_m	x_{m1}	x_{m2}	...	x_{mn}



KARAR AĞAÇLARININ:

- AVANTAJLARI:
- Anlaması ve yorumlaması kolaydır. Kullanılan ağaç yapılar görselleştirilebilir.
- Az oranda bir veri hazırlığına ihtiyaç duyar.

DEZAVANTAJLARI:

Veriyi iyi bir şekilde açıklamayan aşırı karmaşık ağaçlar üretilebilir. Bu durumda ağaç dallanması takip edilemeyebilir.

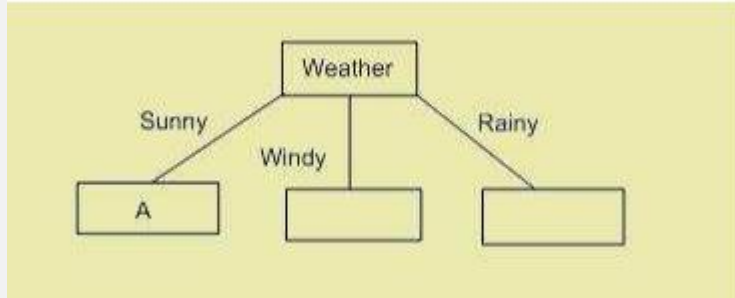


KARAR AĞACI ÖRNEĞİ

- I.Adım: Veri setinden T öğrenme kümesi oluşturulur.

Weekend (Example)	Weather	Parents	Money	Decision (Category)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

- 2.Adım: Veri setindeki en ayırt edici nitelik belirlenir ve ağacın kökü olarak alınır.

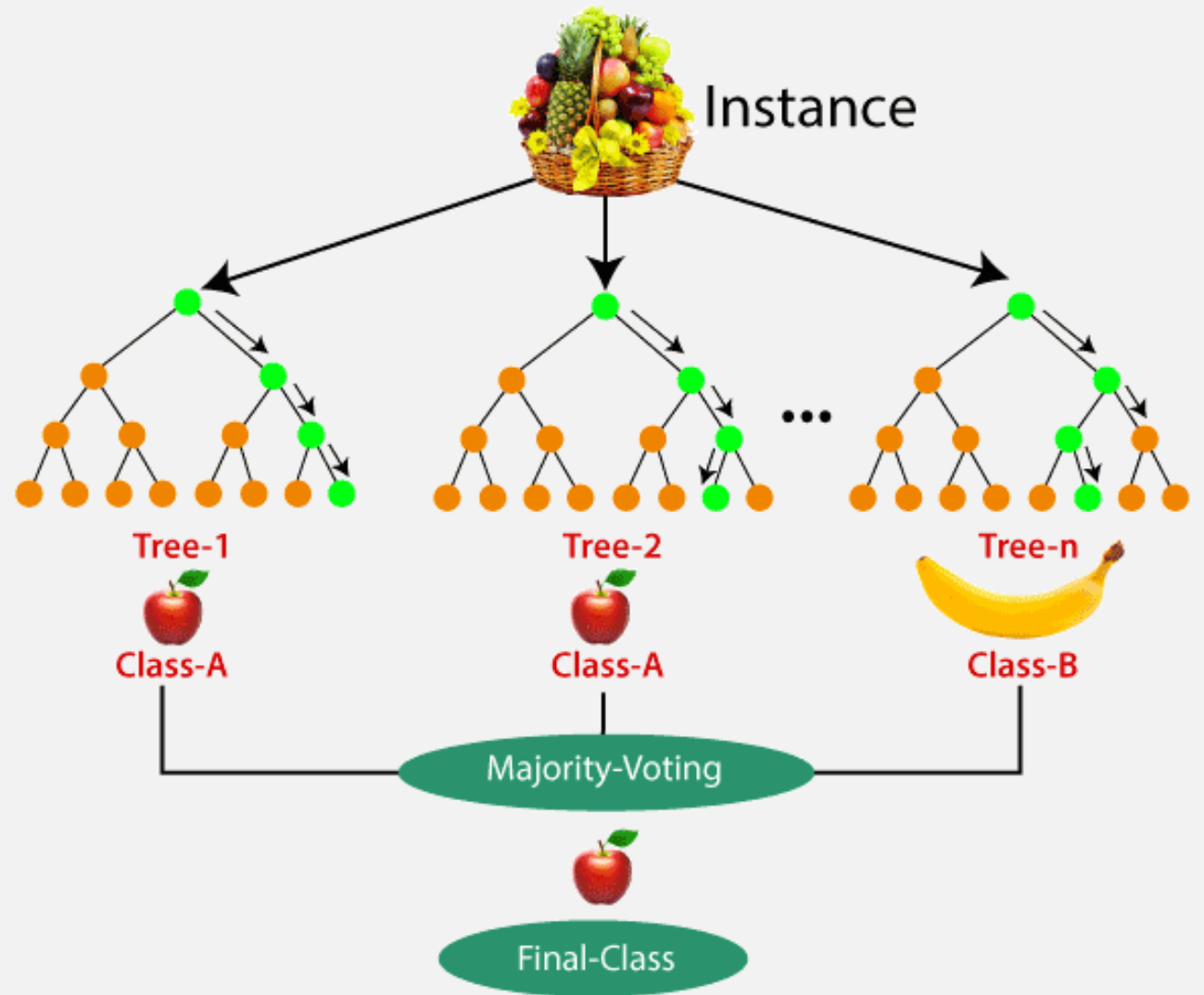


- 3.Adım: Ağacın çocuk düğümü olan A düğümüne ait alt veri kümesi belirlenir.

Weekend (Example)	Weather	Parents	Money	Decision (Category)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W10	Sunny	No	Rich	Tennis

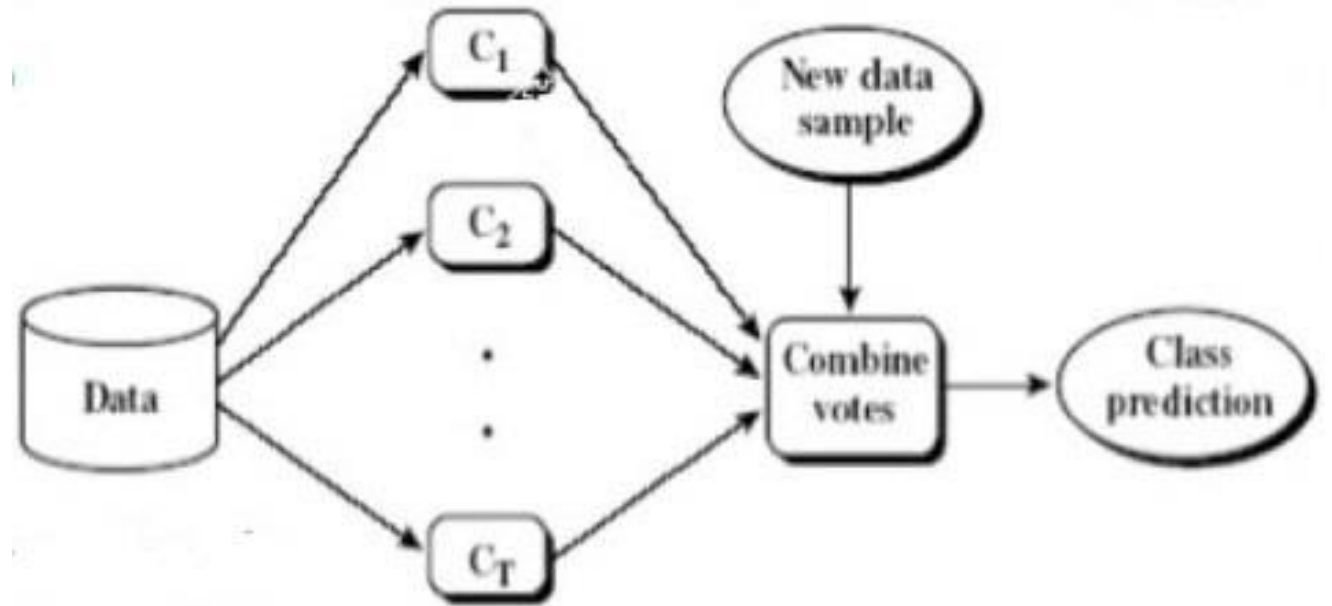
BAGGING

- Bagging: averaging the prediction over a collection of classifiers



BOOSTING

- Boosting: weighted vote with a collection of classifiers



KARAR AĞAÇLARI VE RASTGELE ORMANLAR ARASINDAKİ FARKLAR

- Rastgele ormanlar, karar ağaçlarının bir koleksiyonudur, ancak bazı farklılıklar vardır:



DECISION TREE
VERSUS
RANDOM FOREST

DECISION TREE	RANDOM FOREST
A decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility	An ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class depending on the individual trees
There is a possibility of overfitting	Reduced risk of overfitting
Gives less accurate results	Gives more accurate results
Simpler and easier to understand, interpret and visualize	Comparatively more complex

OVERFITTING

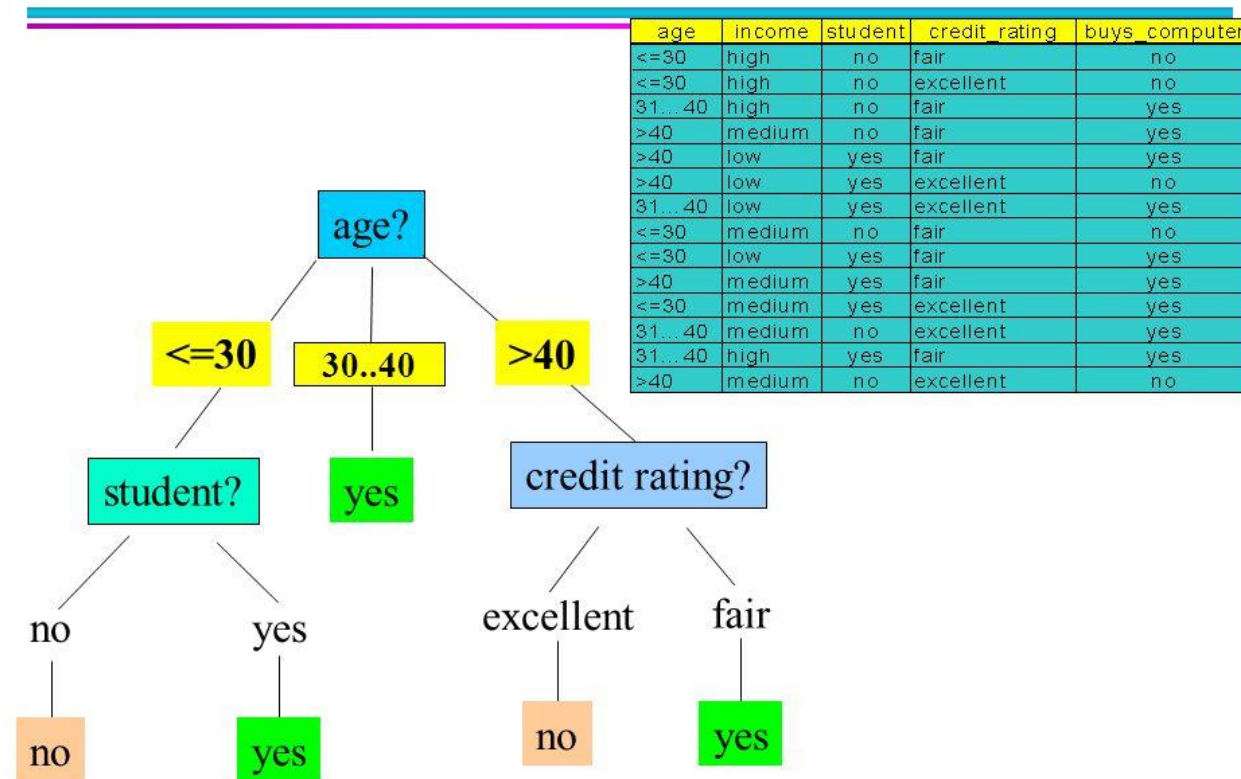
Sistemin aşırı
eğitilmesi sonucunda:

→ Esnekliğini
kaybetmesi

→ Ezberlemeye yakın
bir sonuç oluşturması

→ Gereksiz yere fazla
detay içermesi

Output: A Decision Tree for “*buys_computer*”



BUNLARDAN KAÇINMAK İÇİN; PRUNING(BUDAMA)

1

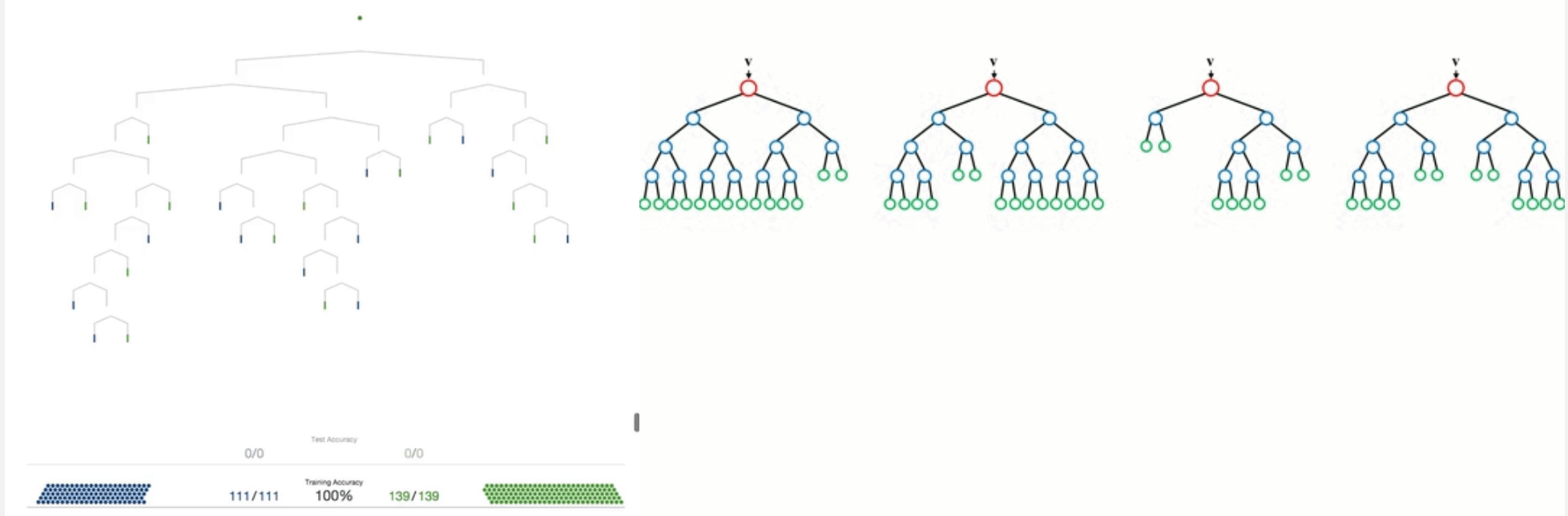
Prepruning: Ağacı
oluştururken budama işlemi.

2

Postpruning: Ağacı
oluşturduktan sonra budama
işlemi.

RANDOM FOREST ALGORİTMASI

- Tek bir ağacın aksine birden fazla ağaç yetiştirilir.
- Sınıflandırma durumunda ormandaki tüm ağaçlar arasında en çok oyu alan seçilir.
- Regresyon durumunda, farklı ağaçların çıktılarının ortalamasını alır.



AVANTAJLARI

- Bu algoritma çok kararlı bir algoritmadır.
- Mevcut makine öğrenmesi sistemlerinin çoğunu oluşturan hem sınıflandırma hem de regresyon problemleri için kullanılabilir.



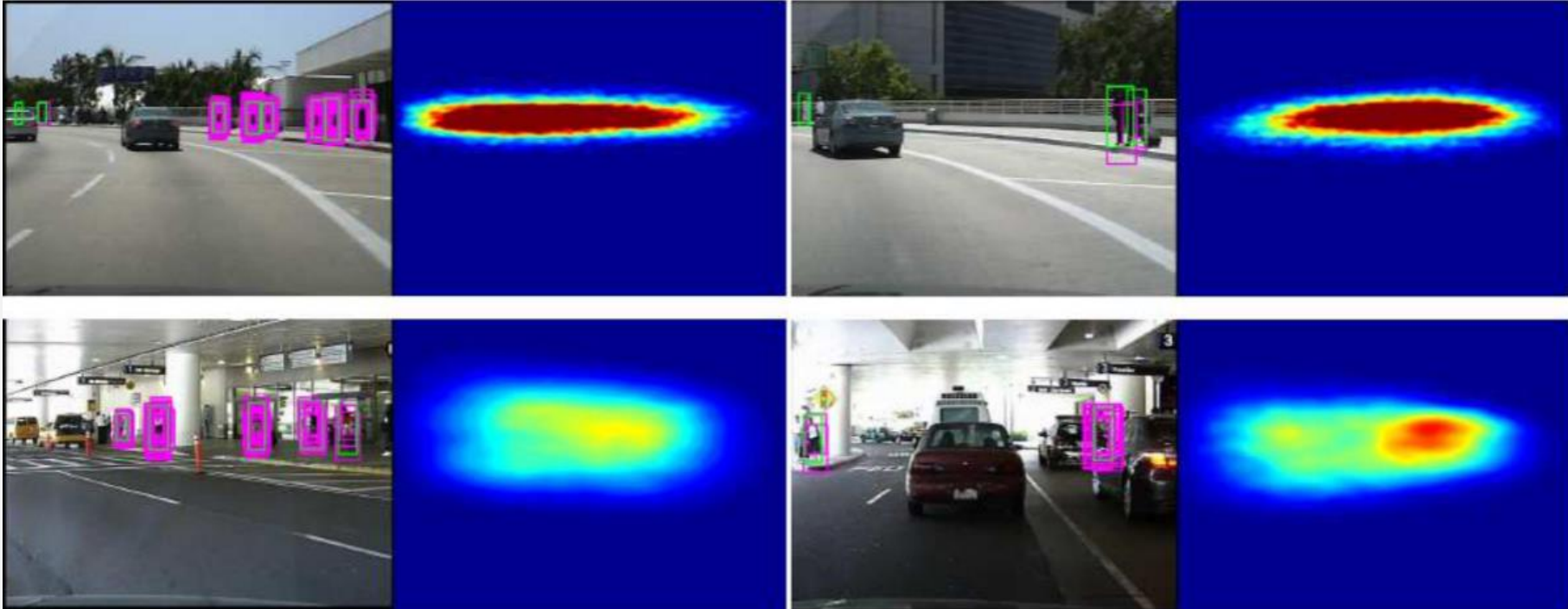
DEZAVANTAJLARI

- Genel olarak, bu algoritmalar hızlı bir şekilde eğitilebilir, ancak bir kez eğitildiklerinde tahminler oluşturmak için oldukça yavaştır.
- Sınıflandırmada iyi ama regresyonda o kadar iyi değildir.



NERELERDE KULLANILIR

- Görüntü sınıflandırması
- Nesne Algılama
- Nesne Takibi
- Kenar algılama
- Anlamsal Bölümleme



NERELERDE KULLANILIR

- İnsan / El Duruşu Tahmini
- 3D Yerelleştirme
- Düşük Seviye Görme
- Yüz İfadesi Tanıma
- Yorumlanabilirlik, düzenlileştirme, sıkıştırma budama ve özellik seçimi



Real-Time Human Pose Recognition in Parts from Single Depth Images

CVPR 2011

Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp,
Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake

Microsoft[®]
Research

Xbox Incubation

RANDOM FOREST ALGORİTMASIYLA E-POSTA SINIFLANDIRMA UYGULAMASI

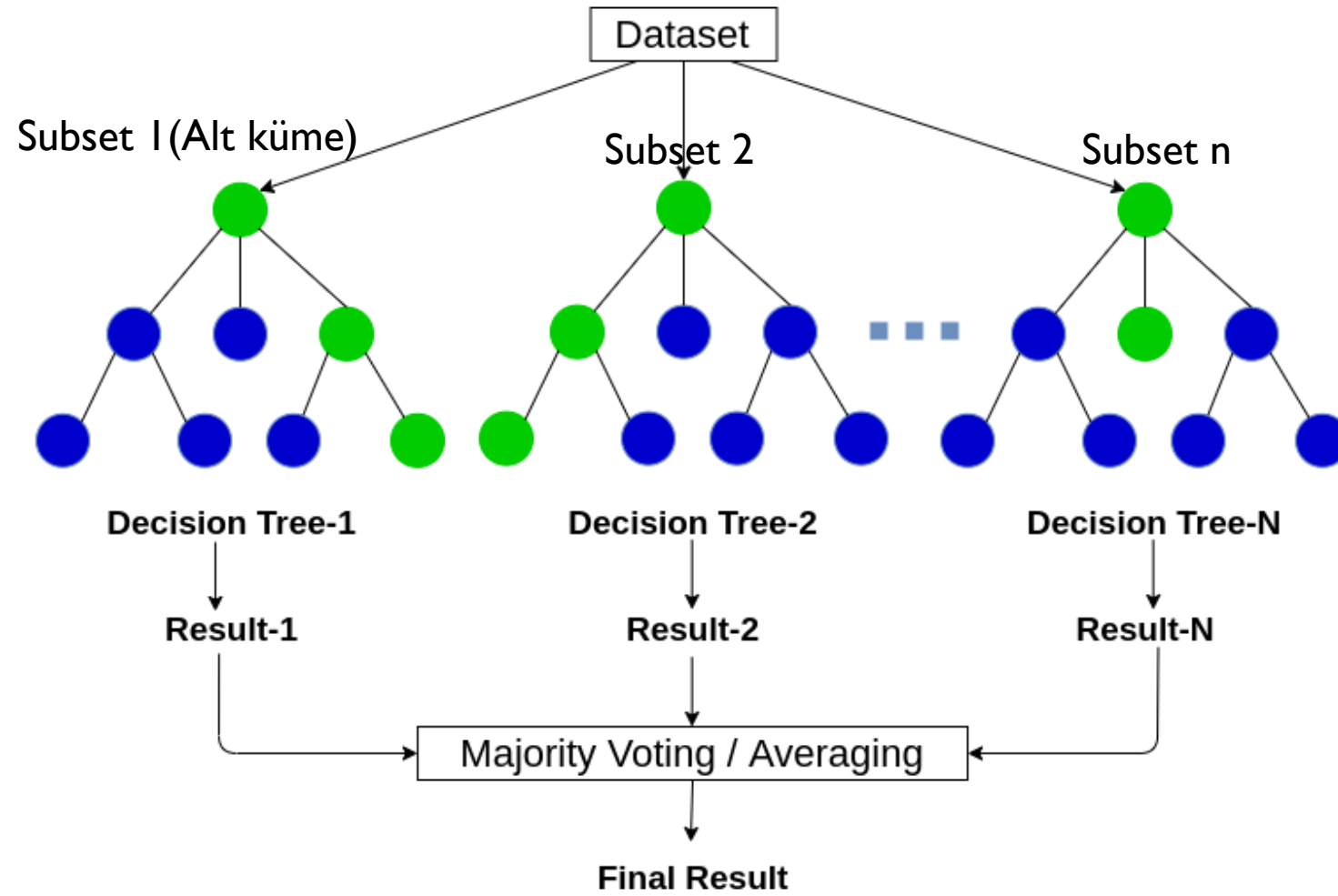


SPAM



SPAM DEĞİL





EĞİTİM DATASI (TRAINING)



MODEL ÖĞRENMESİ



TEST DATASI



TAHMİNLEME (PREDICTION)






OYLAMA

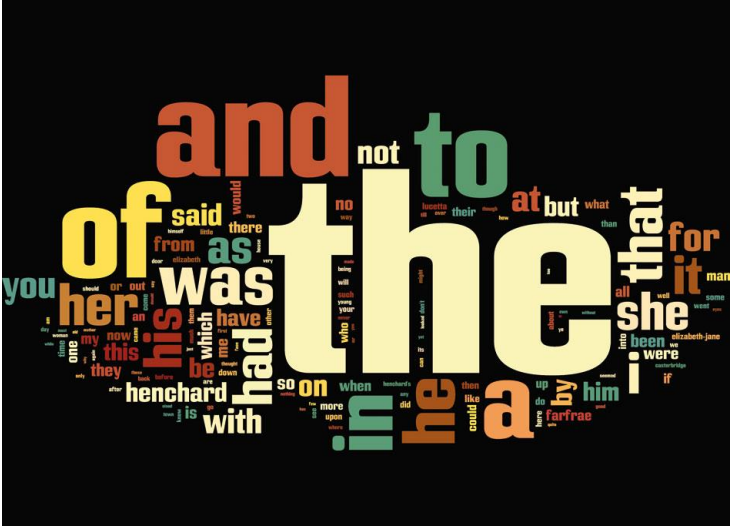


SONUÇ

INPUT MATRIX

kelimeler →	adress	good	course	subject	our	program
	5	1	3	5	0	4
	0	3	0	1	2	1
...						
						
n						
↓						
mail sayısı						

Sample text with Stop Words	Without Stop Words
GeeksforGeeks – A Computer Science Portal for Geeks	GeeksforGeeks , Computer Science, Portal ,Geeks
Can listening be exhausting?	Listening, Exhausting
I like reading, so I read	Like, Reading, read



3-379msg1.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

Subject: job - university utah

linguistic program university utah invite application one-year visit assistant
professor position begin september , 1992 . minimum degree requirement . b . d .
candidate expect teach introductory undergraduate linguistics course course
american english english teach major . propose undergraduate m . . level course
general linguistics sociolinguistic . send letter application , curriculum vita ,
sample publication , three letter reference mauricio mixco , director , linguistic
program , stewart build 213 , university utah , salt lake city , ut 84112 .
further information telephone : 801-581 - 7432 email dipaolo @ anthro . utah . edu
.

spmsgc92.txt - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

Subject: lucky !

congratulation ! ' ve select enter vacation adventure florida ! limit bring
vacation lifetime promote tourism florida ! ' re lucky one ! n't waste wonder ' ll
, ! everyone register qualify ! name keep confidential , never sell list . click
below secure name today ! http : / / -796264397 / travel22 pleasen reply : mailto
: remlst @ 2bmail . co . uk remove list ! great day !
|

Random Forest Algoritmasının Performansı (Confusion Matrix)

Actual	Predicted		
		Negative	Positive
	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

N:Negative(spam)

P:Positive(spam değil)

HESAP

- Accuracy = $(TN+TP)/(TN+TP+FN+FP)$
- Precision = $(TP)/(TP+FP)$
- Recall = $(TP)/(TP+FN)$
- F1 Score = $2 * \frac{Prec * Rec}{Prec + Rec}$