

# YILDIZ TEKNİK ÜNİVERSİTESİ

## BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



### Yapay Zeka Dersi

Prof.Dr. Mehmet Fatih AMASYALI

Ödev No:1

**Ödev Konusu:** Genetik algoritma / Tepe tırmanma ile metin sınıflandırma

**Öğrenci 1 Adı Soyadı:** Buse Nur Pekmezci

**Öğrenci 1 Numara:** 20011001

**Öğrenci 2 Adı Soyadı:** Asude Merve Ekiz

**Öğrenci 2 Numara:** 20011623

Mayıs, 2023

## Ön Bilgiler

Bu çalışmada genetik algoritma ile metin sınıflandırma yapılması amaçlanmıştır. Kodu gerçekleştirmek için Jupyter Notebook ve Python kullanılmıştır. Ayrıca veri işleme ve görselleştirme kolaylaştırmak için numPy, Pandas, matplotlib gibi kütüphanelerden yararlanılmıştır.

5 adet veri seti kullanılmış olup bu veri setleri aşağıdaki gibidir:

1. Amazon Kullanıcı Yorumları:
2. Kültür – Ekonomi Tweetleri:
3. Siyaset – Sağlık Tweetleri
4. Gerçek – Yalan Haber Metinleri
5. 6 Şubat Depremi İle İlgili Yardım Talebi İçeren – İçermeyen Tweetler

İncelenmek üzere 4 adet hiperparametre seçilmiştir. Bu hiperparametreler aşağıdaki gibidir:

1. Popülasyondaki Birey Sayısı
2. Birey Uzunluğu
3. Mutasyon Oranı
4. İterasyon Sayısı

Bahsi geçen hiperparametreler ilk veri setimiz olan Amazon Kullanıcı Yorumları üzerinden incelenecek, bulgularımız ve grafiklerimiz bu veri seti üzerinden yorumlanacaktır. Raporun kalan kısmında diğer veri setlerine dair grafikler de bulunmaktadır.

Sonuçlarda bahsi geçen V1 / V2 kavramlarının açıklanması:

Yukarıda bahsettiğimiz hiperparametrelerin haricinde algoritmayı iyileştirmek adına yeni bir yöntem geliştirdik:

V1 kısmında yeni jenerasyonu yalnızca bir önceki jenerasyondan seçtiğimiz bireyler ile üretiyoruz. V2’de ise yeni jenerasyonu kendinden 2 önceki ve 1 önceki jenerasyonun en iyi bireylerini farklı olanlarda alarak üretiyoruz. Bu sayede oluşacak yeni jenerasyonun daha iyi sonuç vermesini bekliyoruz. (Crossover ve mutasyon işlemleri her iki durumda da gerçekleştirilmektedir.)

Raporun kalan kısmının haritası:

- 1) 5 Farklı Veri Kümesi için Üretilen En İyi Kelime Listeleri  
Burada en iyi kelime listeleri ile birlikte yukarıda bahsettiğimiz V1-V2 karşılaştırması içeren grafikler ve bu grafiklere dair yorumlarımız da bulunmaktadır.
- 2) Hiperparametrelerin Sonuçlara Etkisinin Grafiklerle Gösterilmesi ve Yorumlanması  
Hiperparametrelerin sonuçlara etkisi V1 yöntemiyle üretilen jenerasyonlar üzerinde gösterilecektir.
- 3) Bulgulara Dair Genel Yorumlar

## 1) 5 Farklı Veri Kümesi İçin Üretilen En İyi Kelime Listeleri

### Amazon Kullanıcı Yorumları Veri Seti

#### V1 sonuçları

En iyi birey: ['was', 'just', 'works', 'thing', 'really', 'problem', 'has', 'an', 'been', 'be', 'great', 'problems', 'car', 'well', 'recommend', 'headset', 'off', 'best', 'comfortable', 'happy', 'good', 'thing', 'quality', 'working', 'as', 'few', 'than', 'talk', 'have', 'when', 'what', 'car', 'service', 'item', 'only', 'its', 'recommend', 'never', 'me', 'while', 'did', 'while', 'so', 'car', 'time', 'there', 'be', 'not', 'like', 'worst']

En iyi ikinci birey: ['was', 'just', 'works', 'thing', 'really', 'problem', 'has', 'an', 'been', 'be', 'great', 'problems', 'car', 'well', 'recommend', 'headset', 'off', 'best', 'comfortable', 'happy', 'good', 'thing', 'quality', 'working', 'as', 'few', 'than', 'talk', 'any', 'first', 'problems', 'at', 'really', 'nokia', 'only', 'not', 'have', 'because', 'me', 'last', 'that', 'working', 'time', 'have', 'but', 'there', 'hear', 'also', 'like', 'work']

En iyi üçüncü birey: ['it's', 'just', 'bluetooth', 'hear', 'use', 'service', 'has', 'do', 'best', 'life', 'great', 'any', 'car', 'well', 'recommend', 'am', 'all', 'best', 'from', 'happy', 'good', 'would', 'really', 'working', 'product', 'nice', 'while', 'you', 'one', 'i've', 'get', 'device', 'before', 'from', 'plug', 'ever', 'at', 'not', 'life', 'used', 'everything', 'but', 'hear', 'device', 'cell', 'plug', 'worked', 'poor', 'more', 'last']

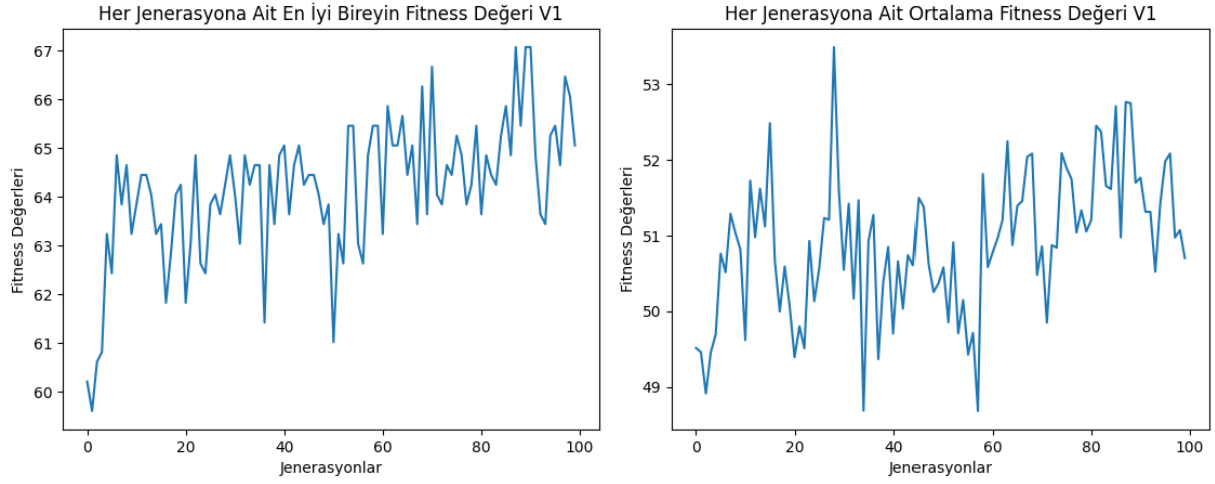
#### V2 sonuçları

En iyi birey: ['any', 'device', 'nokia', 'problem', 'highly', 'best', 'excellent', 'while', 'case', 'these', 'bluetooth', 'get', 'well', 'up', 'good', 'are', 'ear', 'again', 'price', 'happy', 'love', 'but', 'even', 'great', 'reception', 'never', 'only', 'used', 'now', 'nokia', 'much', 'case', 'only', 'bad', 'get', 'bluetooth', 'what', 'an', 'reception', 'while', 'plug', 'be', 'got', 'enough', 'or', 'not', 'service', 'also', 'poor', 'while']

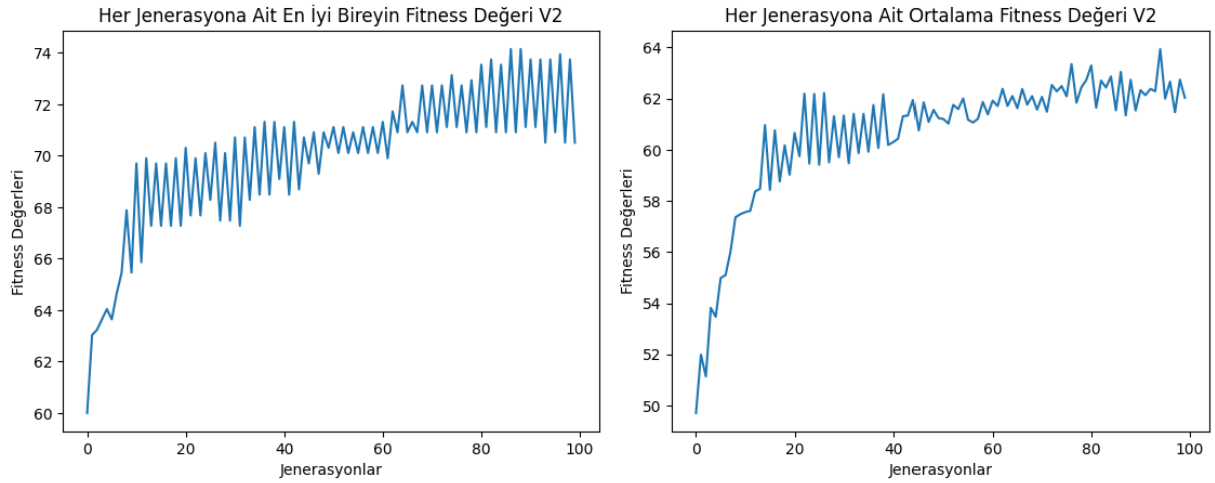
En iyi ikinci birey: ['any', 'device', 'nokia', 'problem', 'highly', 'best', 'excellent', 'while', 'case', 'these', 'bluetooth', 'get', 'well', 'up', 'good', 'are', 'ear', 'again', 'price', 'happy', 'love', 'but', 'even', 'great', 'reception', 'never', 'but', 'worst', 'nice', 'nokia', 'much', 'case', 'much', 'did', 'get', 'bluetooth', 'what', 'an', 'reception', 'while', 'plug', 'be', 'got', 'enough', 'or', 'not', 'service', 'also', 'poor', 'while']

En iyi üçüncü birey: ['any', 'device', 'nokia', 'problem', 'highly', 'best', 'excellent', 'while', 'case', 'these', 'bluetooth', 'get', 'well', 'up', 'good', 'are', 'ear', 'again', 'price', 'happy', 'love', 'but', 'even', 'great', 'reception', 'never', 'but', 'worst', 'nice', 'nokia', 'much', 'case', 'much', 'did', 'get', 'bluetooth', 'what', 'an', 'reception', 'while', 'plug', 'be', 'got', 'enough', 'or', 'not', 'service', 'also', 'poor', 'while']

## V1 grafikleri



## V2 grafikleri



Grafikleri karşılaştırdığımızda yeni jenerasyonu kendinden önceki 2 jenerasyonun en iyi bireyleriyle oluşturduğumuz versiyon olan V2'de, ilk jenerasyonun en iyi bireyinin skoru ile son jenerasyonun en iyi bireyinin skor farkı V1 versiyonuna göre daha fazladır. Yani V2'nin ivmesi V1'den yüksektir. Ayrıca artış daha düzenli, salınım daha azdır. Bu sebeple ata bireylerin en iyilerinin seçilme yönteminin daha iyi bir jenerasyon üretme noktasında faydalı olduğuna kanaat getirdik.

Ortalama fitness değerleri de En İyi Birey değerleri gibi V2'de daha düzenli bir artış göstermiştir.

## Kültür – Ekonomi Tweetleri Veri Seti

### V1 sonuçları

En iyi birey: ['bütün', 'gölü', 'havalimanı', 'daha', 'alan', 'eylül', 'sümela', 'kültür', 'kazandırılacak', 'olduğu', 'in', 'projesi', 'antik', 'nın', 'günü', 'dünyaca', 'muğla', 'insanlar', 'sahipliği', 'dünyanın', 'bodrum', 'adresi', 'hızla', 'yer', 'yolu', 'yeniden', 'eylül', 'bakan', 'değer', 'konusunda', 'çalışanlar', 'turizmine', 'geldi', 'yolları', 'örgütü', 'neden', 'olsun', 'yolu', 'milyon', 'diye', 'lik', 'önceki', 'akdeniz', 'sektörünün', 'yerden', 'olmak', 'alan', 'otel', 'başkenti', 'sadece']

En iyi ikinci birey: ['bütün', 'gölü', 'havalimanı', 'daha', 'alan', 'eylül', 'sümela', 'kültür', 'kazandırılacak', 'olduğu', 'in', 'projesi', 'antik', 'nın', 'günü', 'dünyaca', 'muğla', 'insanlar', 'sahipliği', 'dünyanın', 'bodrum', 'adresi', 'hızla', 'yer', 'yolu', 'yeniden', 'eylül', 'olarak', 'önemli', 'turizmin', 'tatile', 'biz', 'farklı', 'hangi', 'geçen', 'otel', 'yaklaşık', 'havalimanı', 'ab', 'ben', 'müzesi', 'tarafından', 'ağırladı', 'turizmciler', 'dünyada', 'teknoloji', 'arasında', 'olsun', 'birçok', 'birçok']

En iyi üçüncü birey: ['bütün', 'gölü', 'havalimanı', 'daha', 'alan', 'eylül', 'sümela', 'kültür', 'kazandırılacak', 'olduğu', 'in', 'projesi', 'antik', 'nın', 'günü', 'dünyaca', 'muğla', 'insanlar', 'sahipliği', 'dünyanın', 'bodrum', 'adresi', 'hızla', 'yer', 'yolu', 'yeniden', 'eylül', 'bakan', 'önemli', 'sektörünün', 'yolcu', 'desteği', 'yabancı', 'yolları', 'saat', 'artık', 'olsun', 'yolu', 'göl', 'diye', 'örgütü', 'önceki', 'değil', 'sayısı', 'bakanlığı', 'olmak', 'alan', 'daha', 'başkenti', 'i']

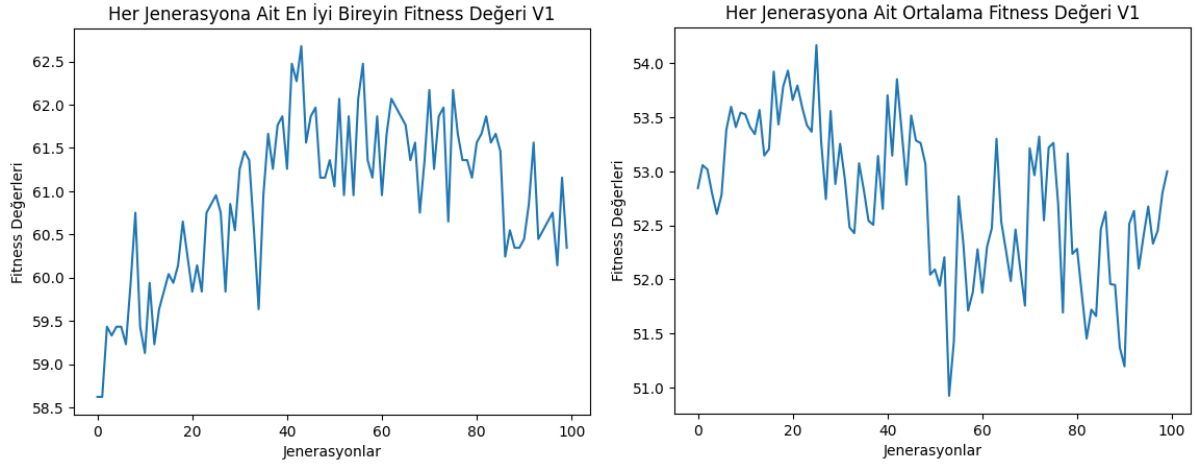
### V2 sonuçları

En iyi birey: ['estetik', 'un', 'tarihi', 'açıldı', 'davet', 'kulesi', 'bağlı', 'dönemi', 'marmaris', 'sümela', 'araya', 'doğal', 'sanat', 'urfa', 'yolcu', 'corona', 'sektörünün', 'gölü', 'bodrum', 'ediyoruz', 'dönemi', 'uçak', 'güzel', 'yeniden', 'bekliyor', 'kapadokya', 'yı', 'yine', 'tanıtım', 'zaman', 'etti', 'ülkenin', 'aylık', 'genel', 'yi', 'işte', 'olarak', 'ticaret', 'alman', 'hava', 'pandemi', 'oteli', 'piknik', 'mavi', 'sayısı', 'tanıtım', 'değil', 'göre', 'yol', 'sektörünün']

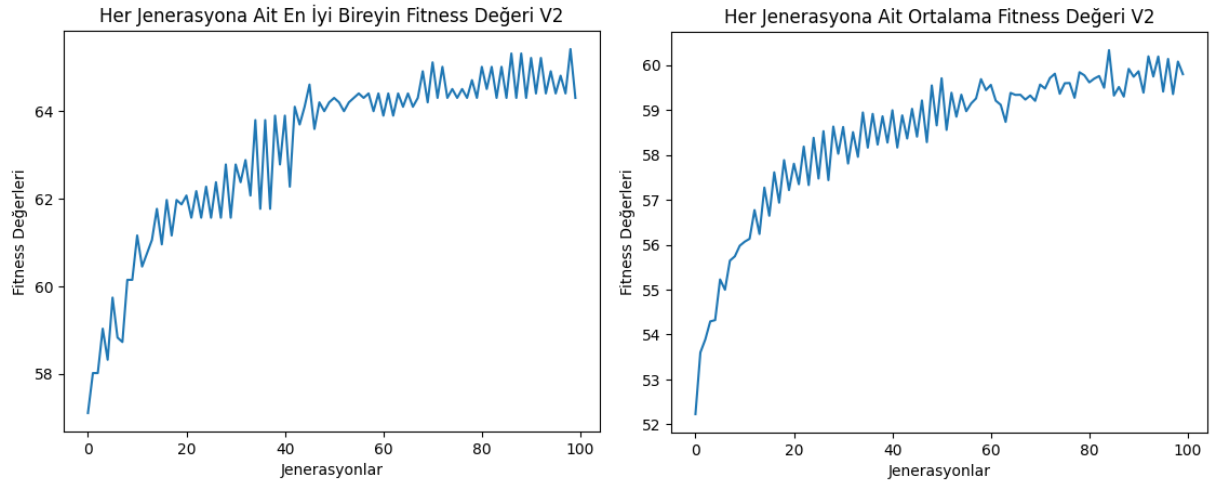
En iyi ikinci birey: ['estetik', 'un', 'tarihi', 'açıldı', 'davet', 'kulesi', 'bağlı', 'dönemi', 'marmaris', 'sümela', 'araya', 'doğal', 'sanat', 'urfa', 'yolcu', 'corona', 'sektörünün', 'gölü', 'bodrum', 'ediyoruz', 'dönemi', 'uçak', 'güzel', 'yeniden', 'bekliyor', 'kapadokya', 'yı', 'yine', 'tanıtım', 'zaman', 'etti', 'ülkenin', 'aylık', 'genel', 'yi', 'işte', 'olarak', 'ticaret', 'alman', 'hava', 'pandemi', 'oteli', 'piknik', 'mavi', 'sayısı', 'tanıtım', 'değil', 'göre', 'yol', 'sektörünün']

En iyi üçüncü birey: ['estetik', 'un', 'tarihi', 'açıldı', 'davet', 'kulesi', 'bağlı', 'dönemi', 'marmaris', 'sümela', 'araya', 'doğal', 'sanat', 'urfa', 'yolcu', 'corona', 'sektörünün', 'gölü', 'bodrum', 'ediyoruz', 'dönemi', 'uçak', 'güzel', 'yeniden', 'bekliyor', 'kapadokya', 'tek', 'yine', 'alman', 'az', 'değerli', 'ayına', 'istiyor', 'otel', 'bursa', 'işte', 'olarak', 'ticaret', 'alman', 'hava', 'pandemi', 'oteli', 'piknik', 'mavi', 'sayısı', 'tanıtım', 'değil', 'göre', 'yol', 'sektörünün']

## V1 grafikleri



## V2 grafikleri



Grafikleri karşılaştırdığımızda V2 versiyonunun V1'e göre daha kararlı bir şekilde artış gösterdiğini gözlemledik. Min – max değerlerinin ise En İyi Bireyin Fitness Değeri grafiklerinde birbirine yakın olduğunu gördük. Ortalama Fitness değerleri ise V2'de V1'e göre daha fazla artış gösterdi.

## Siyaset – Sağlık Tweetleri Veri Seti

### V1 sonuçları

En iyi birey: ['yeni', 'hava', 'sayın', 'var', 'kültür', 'da', 'tatil', 'a', 'çok', 'türkiye', 'yok', 'avrupa', 'da', 'ersoy', 'antalya', 'güzel', 'ya', 'türkiye', 'dünya', 'alanya', 'tan', 'sonra', 'ile', 'bakan', 'belediye', 'test', 'değil', 'dikkat', 'covid', 'sağlıklı', 'nin', 'şu', 'pkk', 'nin', 'ama', 'ya', 'covid', 'bu', 'devam', 'gün', 'olan', 'sayısı', 'başladı', 'hayırlı', 'sağlık', 'her', 'belediye', 'salgını', 'isim', 'o']

En iyi ikinci birey: ['yeni', 'isim', 'güzel', 'hava', 'ya', 'da', 'tatil', 'a', 'açıklama', 'türkiye', 'yok', 'süreci', 'nin', 'ersoy', 'antalya', 'gibi', 'ya', 'ki', 'gibi', 'alanya', 'test', 'sonra', 'ile', 'bakan', 'belediye', 'isim', 'iyi', 'gün', 'sağlık', 'sağlıklı', 'nin', 'şu', 'pkk', 'covid', 'olacak', 'test', 'en', 'dünya', 'yaparken', 'sadece', 'sonra', 'ama', 'pandemi', 'olduğu', 'bakanı', 'turist', 'güzel', 'sayısı', 'açıkladı', 'tüm']

En iyi üçüncü birey: ['yeni', 'isim', 'güzel', 'hava', 'ya', 'da', 'tatil', 'a', 'açıklama', 'türkiye', 'yok', 'süreci', 'nin', 'ersoy', 'antalya', 'gibi', 'ya', 'ki', 'gibi', 'alanya', 'test', 'sonra', 'ile', 'bakan', 'belediye', 'isim', 'iyi', 'gün', 'sağlık', 'sağlıklı', 'nin', 'şu', 'pkk', 'covid', 'olacak', 'test', 'en', 'dünya', 'yaparken', 'sadece', 'sonra', 'ama', 'pandemi', 'olduğu', 'bakanı', 'turist', 'güzel', 'sayısı', 'açıkladı', 'tüm']

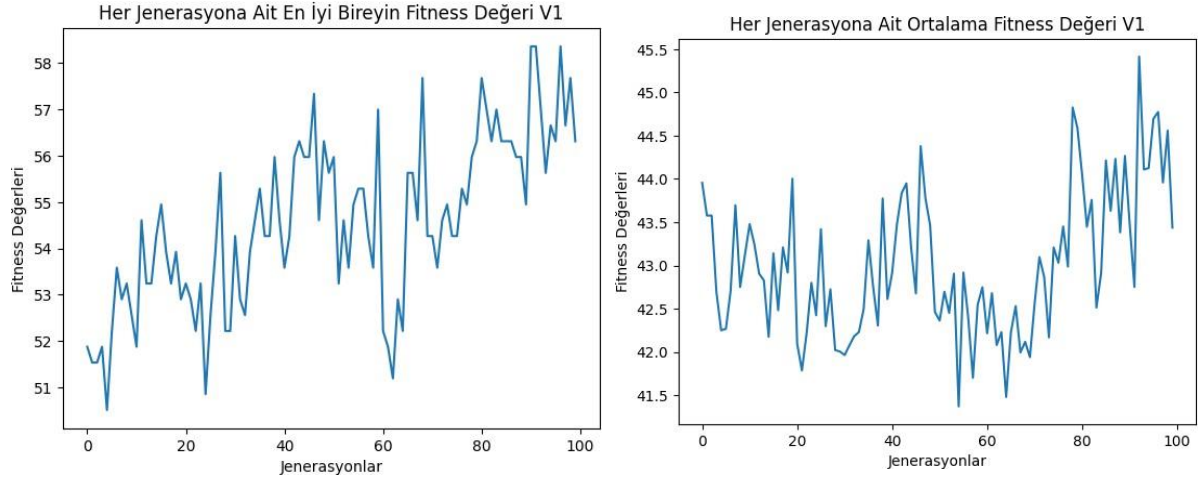
### V2 sonuçları

En iyi birey: ['güzel', 'olur', 'yerine', 'değil', 'açıklama', 'gün', 'var', 'büyük', 'ne', 'alanya', 'nin', 'hava', 'sayın', 'iyi', 'turist', 'yaptı', 'ersoy', 'antalya', 'turist', 'seyahat', 'bir', 'daha', 'da', 'adil', 'karantina', 'sadece', 'o', 'koronavirüs', 'dikkat', 'kadar', 'olacak', 'hayırlı', 'güzel', 'de', 'sağlıklı', 'karşı', 'güzel', 'şu', 'vaka', 'tan', 'için', 'olan', 'antalya', 'hem', 'sağlıklı', 'salgını', 'vaka', 'bakanı', 'vaka', 'huzurlu']

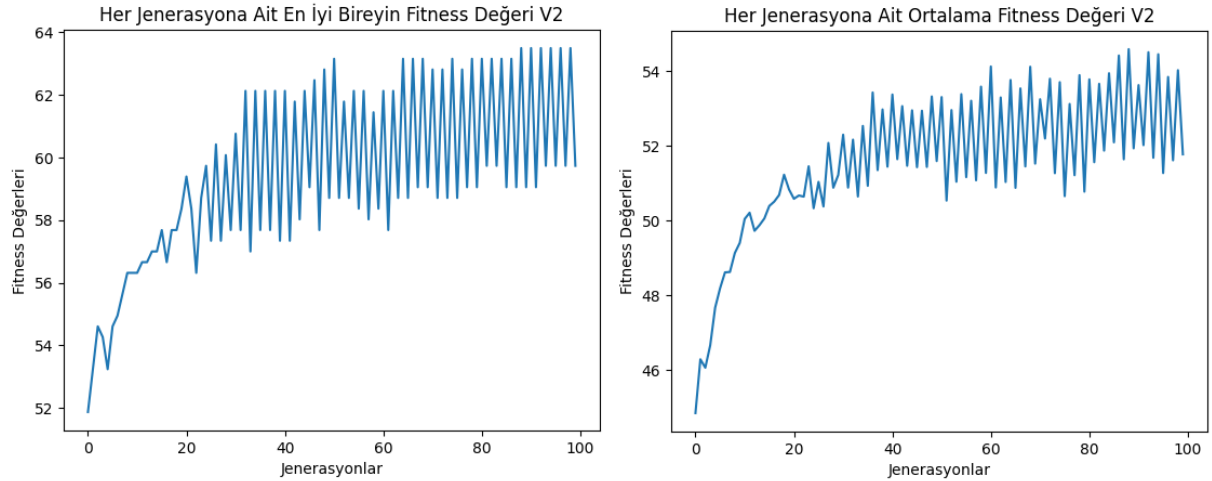
En iyi ikinci birey: ['güzel', 'olur', 'yerine', 'değil', 'açıklama', 'gün', 'var', 'büyük', 'ne', 'alanya', 'nin', 'hava', 'sayın', 'iyi', 'turist', 'yaptı', 'ersoy', 'antalya', 'turist', 'seyahat', 'bir', 'daha', 'da', 'adil', 'karantina', 'sadece', 'o', 'koronavirüs', 'hava', 'kadar', 'olacak', 'hayırlı', 'güzel', 'de', 'sağlıklı', 'karşı', 'güzel', 'şu', 'vaka', 'tan', 'için', 'olan', 'antalya', 'hem', 'sağlıklı', 'salgını', 'vaka', 'bakanı', 'vaka', 'huzurlu']

En iyi üçüncü birey: ['güzel', 'olur', 'yerine', 'değil', 'açıklama', 'gün', 'var', 'büyük', 'ne', 'alanya', 'nin', 'hava', 'sayın', 'iyi', 'turist', 'yaptı', 'ersoy', 'antalya', 'turist', 'seyahat', 'bir', 'daha', 'da', 'adil', 'karantina', 'sadece', 'o', 'koronavirüs', 'dikkat', 'kadar', 'olacak', 'hayırlı', 'güzel', 'de', 'sağlıklı', 'karşı', 'güzel', 'şu', 'vaka', 'tan', 'için', 'olan', 'antalya', 'hem', 'sağlıklı', 'salgını', 'vaka', 'bakanı', 'vaka', 'huzurlu']

## V1 grafikleri



## V2 grafikleri



Sonuçları karşılaştırdığımızda En İyi Birey Fitness Değeri grafiklerinden V2'nin ilk ve son jenerasyon bireyi arasındaki fark V1'e göre daha yüksektir. Dolayısıyla V2 versiyonunda ivmenin daha yüksek olduğundan bahsedilebilir. Ortalama Fitness değerleri grafiklerine baktığımızda da V1 versiyonunda ilk jenerasyondan son jenerasyona yaklaşık 1 birimlik bir düşüş söz konusu iken V2'de net bir yükseliş söz konusudur. Yani yine V2'nin daha iyi bir sonuç verdiği söylenebilir.



## Gerçek – Yalan Haber Metinleri Veri Seti

### V1 sonuçları

En iyi birey: ['vietnam', 'km', 'holding', 'reuters.', 'slightly', 'qatar', 'tuesday.', 'task', 'consider', 'nominations', 'committee.', 'reaching', 'necessary', 'moving', 'indiana', 'mexico', 'itself', 'arrest', 'break,', 'water', 'divided', 'concern', 'indictment', 'clashes', 'low', 'mainstream', 'dismissed', 'chapter', 'damn', 'religious', 'klux', 'hasn', 'accusation', 'chomsky', 'corruption', 'knows', 'golf', 'atta', 'flawed', 'patton', 'occur', 'perpetrators', 'party.', 'spanish', 'alien', 'challenges', 'wall,', 'blacks', 'tactics', 'reluctant']

En iyi ikinci birey: ['seek', 'agency,', 'counter-terrorism', 'reuters.', 'resumed', 'match', 'tuesday.', 'phase', 'current', 'declare', 'touching', 'designer', 'discussions', 'protections', 'wanting', 'backlash', 'ronald', 'develop', 'stated', 'water', 'visit', 'potential', 'research', 'largely', 'effective', 'mass', 'riyadh', 'brotherhood', 'hacking', 'serial', 'obama,', 'front-runner', 'illegals', 'man.', 'communication', '1', 'picked', 'narrative', 'enforcement', 'maryland', 'reducing', 'isn', 'delay', 'sovereign', 'records.', 'ridiculous', 'admitted', 'egyptian', 'attacked', 'christopher']

En iyi üçüncü birey: ['seek', 'per', 'associates', 'reuters.', 'resumed', 'directed', 'tuesday.', 'phase', 'current', 'broadcast', 'half', 'london', 'cnn,', '2016.', 'proud', 'retaliation', 'probe,', 'platform', 'privacy', 'dhs', 'visit', 'cuban', 'allegedly', 'path', 'sought', 'mass', 'uncertainty', 'push', 'cbs', '20.', 'environmental', 'alabama', 'annually', 'guests', 'words,', 'knows', 'partnership', 'weekend,', 'reaching', 'chances', 'rakhine', 'johnson', 'committees', 'purge', 'truly', 'politician', 'polling', 'programs', 'sad', 'refuse']

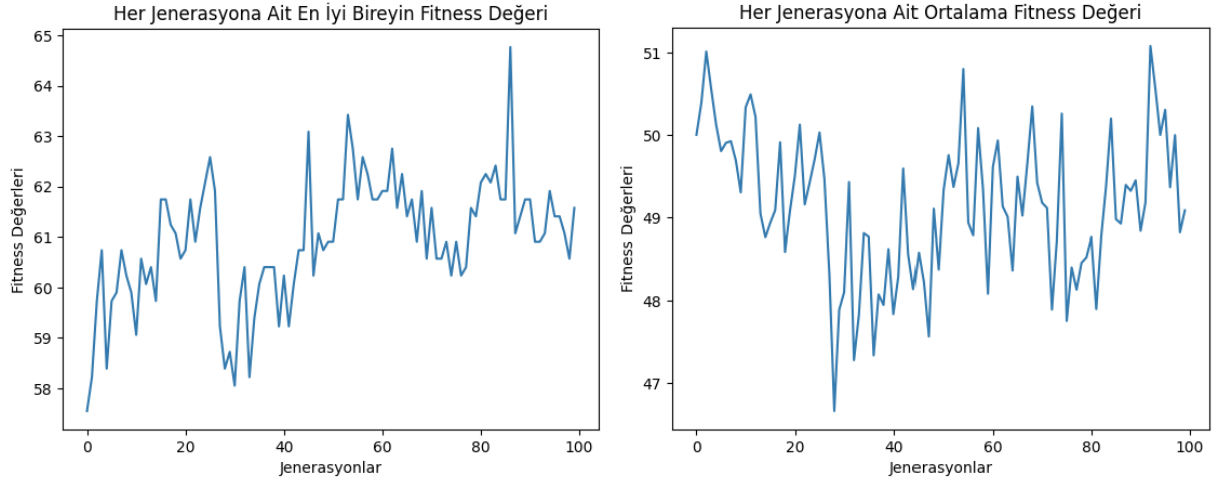
### V2 sonuçları

En iyi birey: ['riina', 'watchdog', 'kurdish', 'killer', 'walk', 'role', 'documents', 'chris', 'seeking', 'lifting', 'if', 'although', 'cuba', 'credibility', 'month,', 'steve', 'search', 'arabia,', 'presumptive', 'equity', 'projected', 'rule', 'stability', 'appeal', 'allowing', 'witnessed', 'kids', 'tells', 'user', 'colleagues', 'be.', 'propaganda', 'smart', 'live', 'talked', 'more:', 'constitution', 'reason', 'currency', 'defending', 'facing', 'reminding', 'head,', 'challenge,', 'residence', 'brought', 'washington.', 'themselves', 'actions', 'uncomfortable']

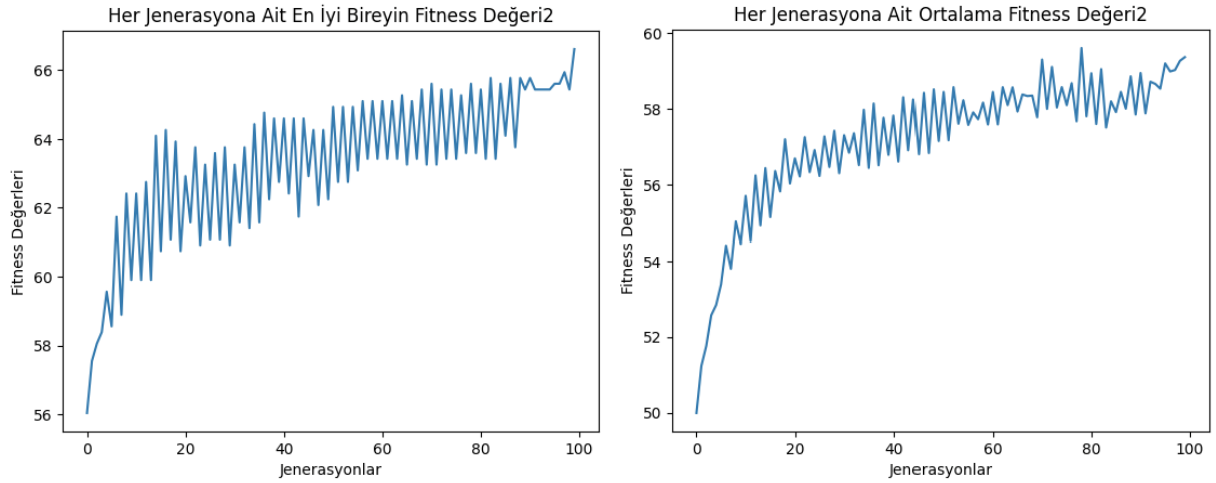
En iyi ikinci birey: ['renzi', 'market', 'points.', 'possibility', 'primary', 'dissent', 'capital', 'el', 'passed', 'union', 'booed', 'oct.', 'counsel', 'interference', 'hospital', 'east', 'standoff', 'combination', 'wouldn', 'moscow', 'assistance', 'income', 'differences', 'kurdish', 'itself,', 'sell', 'target', 'explain', 'rebel', 'hell', 'correct', 'interested', 'race', 'dropped', 'kids', 'worker', 'time.', 'feeling', 'heard', 'acr', 'refugees', 'well,', 'monetary', 'mcauliffe', 'therefore,', 'dash', 'and,', 'you.', 'changed', 'night,']

En iyi üçüncü birey: ['renzi', 'market', 'murder.', 'possibility', 'primary', 'dissent', 'capital', 'dissent', 'passed', 'union', 'violate', 'countries', 'counsel', 'interference', 'hospital', 'east', 'granted', 'murderous', 'bashar', 'moscow', 'assistance', 'credit', 'accepted', 'hospital', 'itself,', 'married', 'problems', 'explain', 'pushing', 'joined', 'tonight', 'footage', 'florida.', 'sheet', 'disrespect', 'worker', 'time.', 'replace', 'violate', 'acr', 'worse', 'well,', 'bail', 'germany', 'to,', 'targeting', 'indeed,', 'focus', 'quietly', 'whoever']

## V1 grafikleri



## V2 grafikleri



Bu veri setinde ise En İyi Birey Fitness Değeri grafikleri karşılaştırıldığında tepe noktaların değeri birbirine yakındır. Ancak V2, V1'e göre daha kararlı bir artış göstermiştir ve artış ivmesi daha büyüktür diyebiliriz.

## 6 Şubat Depremi ile İlgili Yardım Talebi İçeren – İçermeyen Tweetler Veri Seti

### V1 sonuçları

En iyi birey: ['yardım', 'no', 'evler', 'çocuk', 'mazot', 'no:', 'iç', 'battaniye', 'yardımı', 'katına', 'olanlar', 'kahramanmaraş', 'da', 'desteği', 'kadın', 'cumhuriyet', 'aynı', 'çocuk', 'çadır', 'çadır', 'desteği', 'ped', 'caddesi', 'acil', 'yardim', 'teyitli', '20', 'ki', 'destek', 'olur', 'küçük', 'kadınlar', '!', 'mendil', 'yardımı', 'adıyaman', 'ulaşmamış', 'olmaz!', 'bugün', 'gaziantep', 'için', 'bunun', 'küçük', 'ile', 'olur', '20', 'sel', 'kadınlar', '@depremdairesi', 'olanlar']

En iyi ikinci birey: ['yardım', 'no', 'evler', 'çocuk', 'mazot', 'no:', 'iç', 'battaniye', 'yardımı', 'katına', 'olanlar', 'kahramanmaraş', 'da', 'desteği', 'kadın', 'cumhuriyet', 'aynı', 'çocuk', 'çadır', 'çadır', 'desteği', 'ped', 'caddesi', 'acil', 'yardim', 'teyitli', 'yemek', 'powerbank', 'soğuk', 'ped', 'yeni', 'katına', 'meydana', 'bezi', 'kömür', 'sonrası', 'yaşlı', 'barınma', 'mı', 'gün', '30', 'sonrası', 'de', 'hiç', 'malatya', 'hasan', 'kadın', 'o', 'özellikle']

En iyi üçüncü birey: ['mazot', 'gidip', 'evler', 'çocuk', 'mazot', 'ısıtıcı', 'gelen', 'battaniye', 'yardımı', '96', 'kahramanmaraş', 'iletişim', 'no', 'lazım', 'kadın', 'cumhuriyet', 'kahramanmaraş', 'en', 'iç', 'çadır', 'mahallesi', 'defne', '10', 'acil', 'mazot', 'herkes', '@depremdairesi', 'en', 'battaniye', 'bunun', 'gibi', '@depremdairesi', 'sonrası', 'mı', '4', 'ben', 'gaziantep', 'termal', 'şey', 'çocuk', 'ben', 'gelen', 'bezi', 'odun', 'ile', 'yeni', '10', 'termal', 'termal', 'kuru']

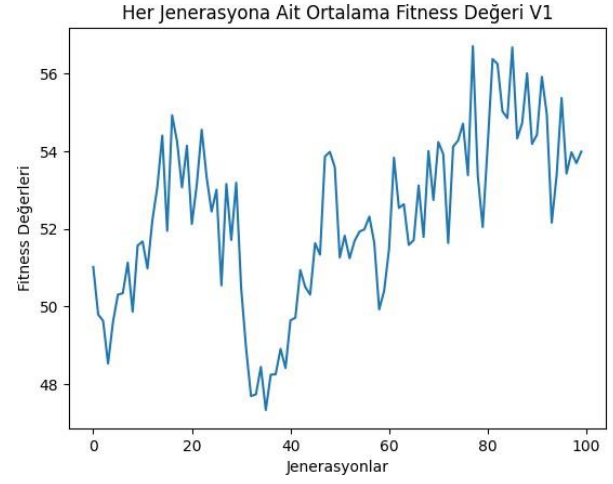
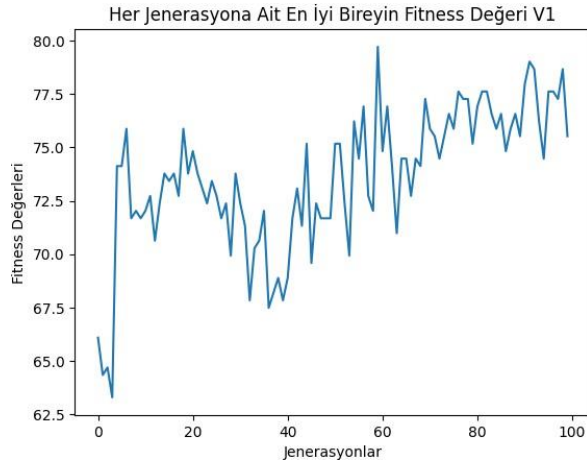
### V2 sonuçları

En iyi birey: ['yardim', 'gelen', 'olan', '16', 'no:', 'bezi', 'arkadaşlar', 'iç', 'hatay', 'yaşlı', 'var.', 'yardımı', 'ahmet', 'no', 'el', 'battaniye', 'ihtiyacı', 'az', 'cumhuriyet', '39', 'teyitli', 'olmaz!', 'lazim', 'kötü', 'bezi', 'acilen', 'çantası', 'gidip', 'olmaz!', 'ki', 'yangın', 'mı', 'meydana', 'gün', 'soğuk', 'mahallesi', 'bu', 'gelen', 'ile', 'ben', '0530', '3', 'pazarcık', 'daha', 'el', '44', 'evler', 'acilen', 'acilen', 'varsa']

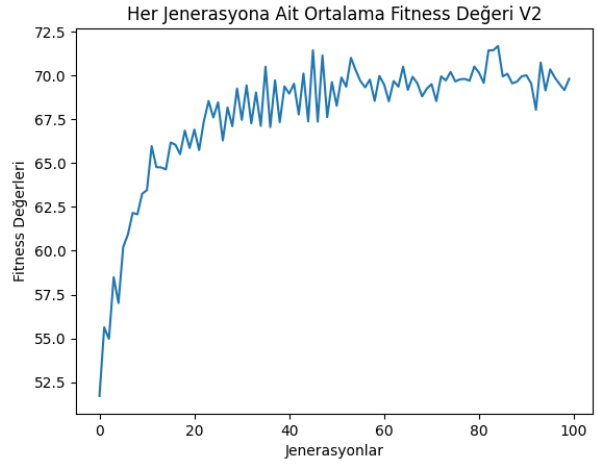
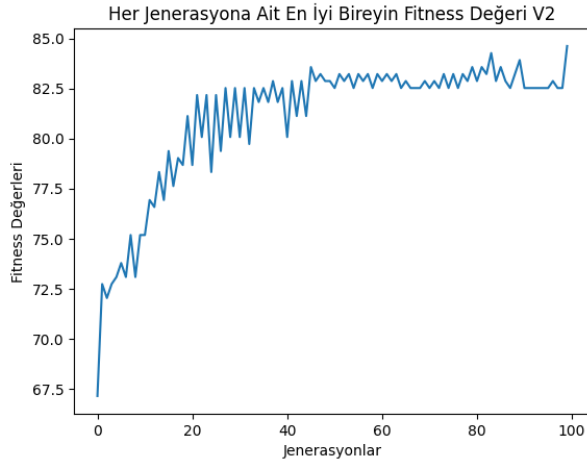
En iyi ikinci birey: ['yardim', 'gelen', 'olan', '16', 'no:', 'bezi', 'arkadaşlar', 'iç', 'hatay', 'yaşlı', 'var.', 'yardımı', 'ahmet', 'no', 'el', 'battaniye', 'ihtiyacı', 'az', 'cumhuriyet', '39', 'teyitli', 'olmaz!', 'lazim', 'kötü', 'bezi', 'acilen', 'çantası', 'gidip', 'malzemeleri', 'ki', 'yangın', 'mı', 'meydana', 'gün', 'soğuk', 'özellikle', 'bu', 'gelen', 'ile', 'ben', 'olunca', '3', 'pazarcık', 'olmaz!', 'el', 'çantası', 'köy', 'acilen', 'acilen', 'varsa']

En iyi üçüncü birey: ['yardim', 'gelen', 'olan', '16', 'no:', 'bezi', 'arkadaşlar', 'iç', 'hatay', 'yaşlı', 'var.', 'yardımı', 'ahmet', 'no', 'el', 'battaniye', 'ihtiyacı', 'az', 'cumhuriyet', '39', 'teyitli', 'olmaz!', 'lazim', 'kötü', 'bezi', 'acilen', 'çantası', 'milyon', 'malzemeleri', 'ki', 'yangın', 'mı', 'meydana', 'gün', 'soğuk', 'kişiler', 'kadar', 'olanlar', 'milyon', 'pazarcık', 'ile', 'mah', 'pazarcık', 'daha', 'el', '3', '/', 'acilen', '#deprem', 'bi']

## V1 grafikleri



## V2 grafikleri



Bu veri seti en yüksek skorları gördüğümüz veri seti oldu. Yardım çağrısında kullanılan kelimeler günlük konuşma dilinden ayrıldığı için bu veri setindeki başarının yüksek olduğunu düşünüyoruz. Ortalama Fitness Değeri grafikleri karşılaştırıldığında V2'nin ivmesi, V1'e göre çok daha yüksektir. Ayrıca genel olarak başarı V2'de, V1'e göre artmıştır ve daha kararlı bir grafik gözlemlenir.

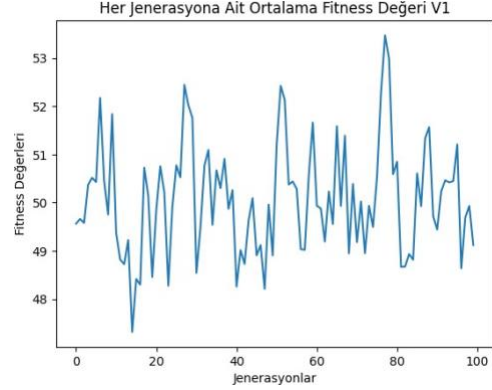
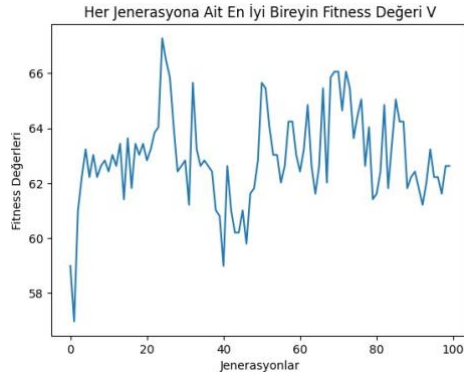
## 2 ) Hiperparametrelerin Sonuçlara Etkisinin Grafikler İle Gösterimi

Hiperparametrelerin etkisi Amazon Kullanıcı Yorumları veri seti üzerinden gösterilecektir:

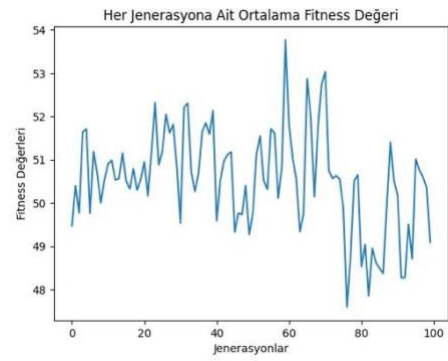
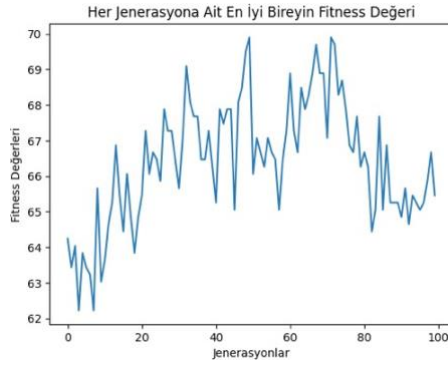
### 1) Popülasyondaki Birey Sayısı değişiminin etkisi

Default olan değerler:

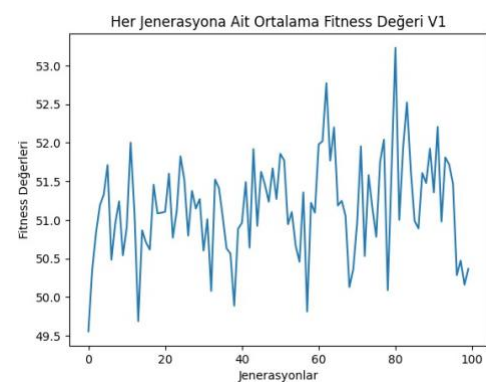
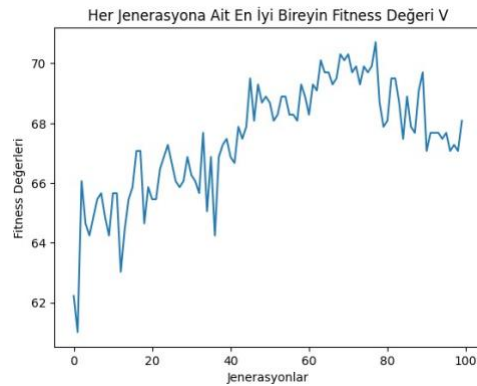
- Birey Uzunluğu: 50
  - Mutasyona uğrayacak birey oranı: 0.4
  - Mutasyona uğrayacak gen oranı: 0.5
- a) Popülasyondaki birey sayısı: 50



### b) Popülasyondaki birey sayısı: 100



### c) Popülasyondaki birey sayısı:200



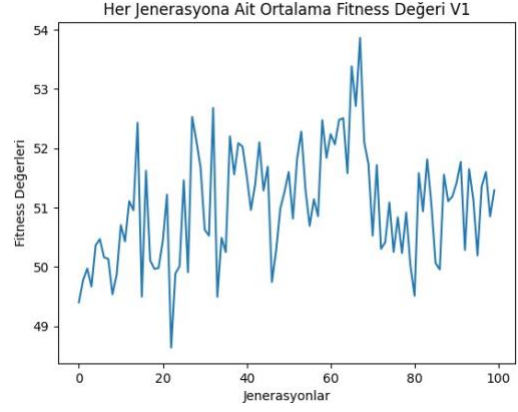
Birey sayısı arttıkça olasılık sayısı da arttığı için en iyi bireyin fitness score'u artmıştır. Belli bir optimum değerden sonra ise pek bir değişiklik olmamıştır. Ortalama fitness değerinde ise iyi bireylerin sayısındaki artış kadar kötü bireylerin sayısında da artış olacağı için kayda değer bir değişiklik gözlenmemiştir.

## 2) Birey uzunluđu deđiřiminin etkisi

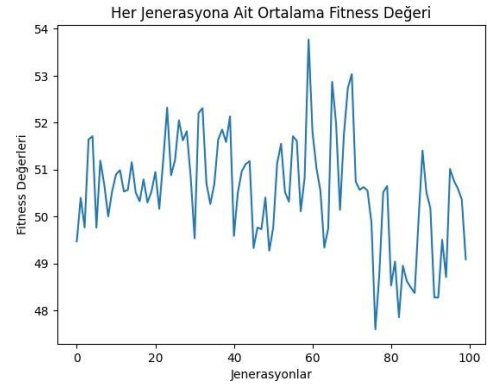
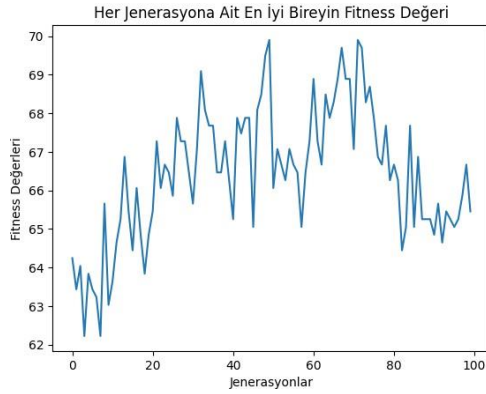
Default olan deđerler:

- Popölasyondaki birey sayısı: 100
- Mutasyona uğrayacak birey oranı: 0.4
- Mutasyona uğrayacak gen oranı: 0.5

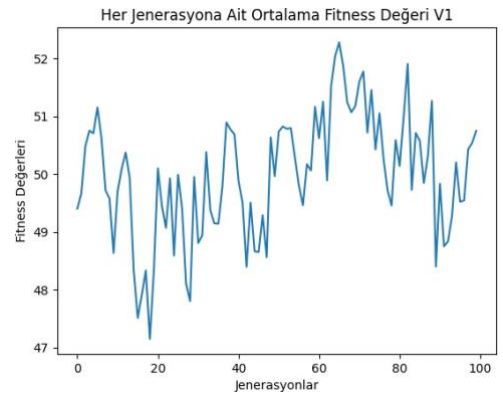
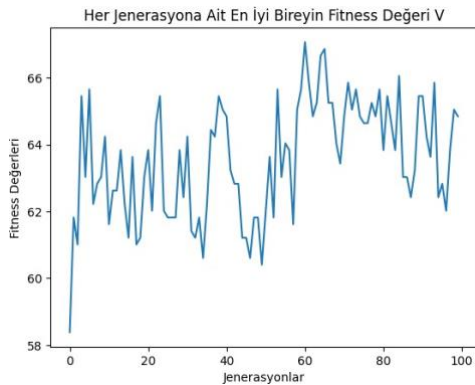
### a) Birey uzunluđu: 20



### b) Birey Uzunluđu:50



### c) Birey Uzunluđu:100



Bu kısımda birey uzunluđu arttıkça grafiklerin deđerlerinde düşüş gözlemledik. Bunun sebebi olarak kullandığımız Amazon kullanıcı yorumları veri setinin kelimelerinin çoğunluğunun olumlu ya da olumsuz duygu içermemesi olduğunu düşünmekteyiz. Örneğin I, you, we, they gibi zamirler yoğunluktadır.

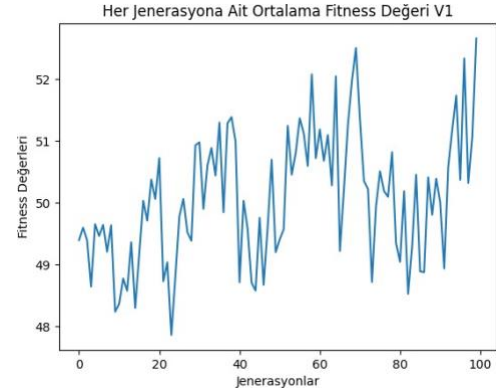
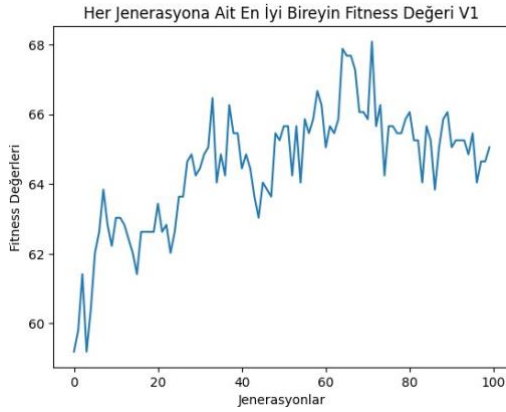
### 3) Mutasyon oranının etkisi

#### 3.1) Mutasyona uğrayacak gen oranı değişiminin etkisi

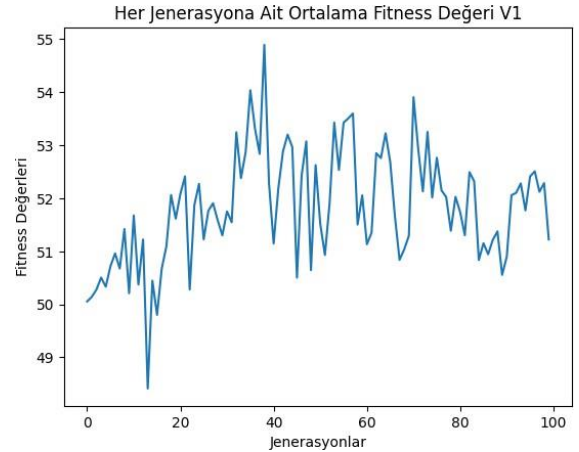
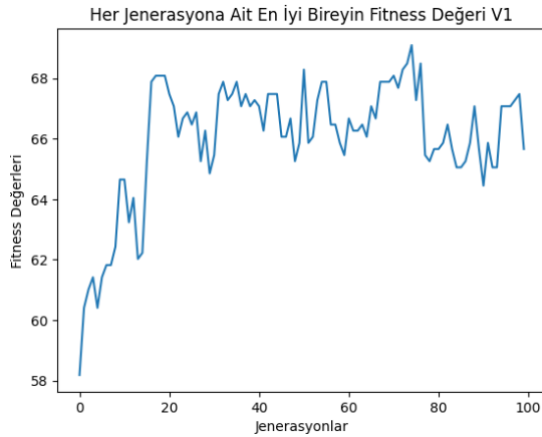
Default olan değerler:

- Popülasyondaki birey sayısı: 100
- Birey uzunluğu: 50
- Mutasyona uğrayacak birey oranı: 0.4

#### a) Mutasyona uğrayacak gen oranı: %25



#### b) Mutasyona uğrayacak gen oranı: %75



En iyi birey grafiklerine baktığımızda mutasyona uğrayacak gen oranı %25 olduğunda artışın yavaş yavaş olduğunu, %75 olduğunda ise ani bir yükselişten sonra aynı değerler arasında salınım yaptığını gözlemlemekteyiz. Bu sebeple mutasyon oranını başta yüksek tutup sonradan düşürmenin daha iyi bir jenerasyon üretmeye katkısı olabileceğini düşünmekteyiz.

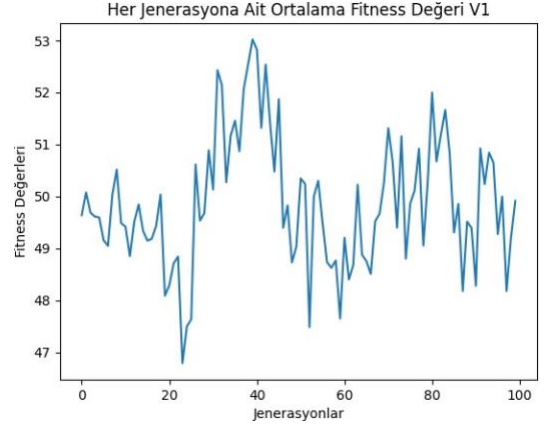
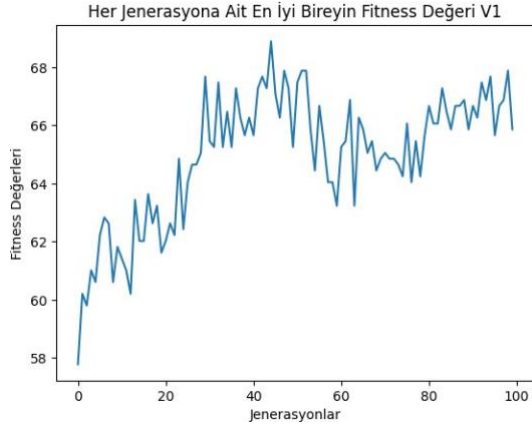
Ortalama fitness değerlerine baktığımızda ise her iki grafiğin en yüksek ortalama değerininin mutasyon oranının yüksek olduğu grafikte daha fazla olduğunu, ilk ve son jenerasyonlar arasındaki farkın ise mutasyon oranının düşük olduğu grafikte daha fazla olduğunu söyleyebiliriz.

### 3.2) Mutasyona uğrayacak birey oranı değişiminin etkisi

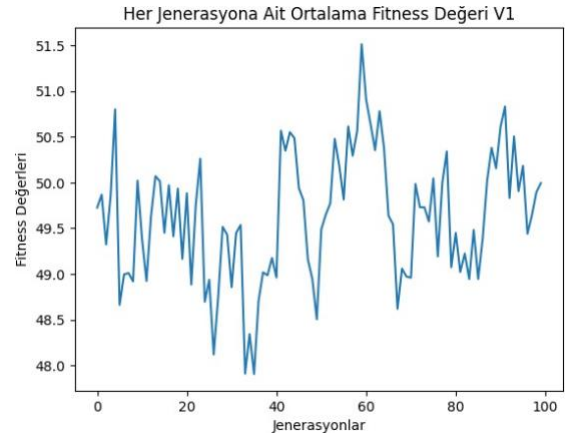
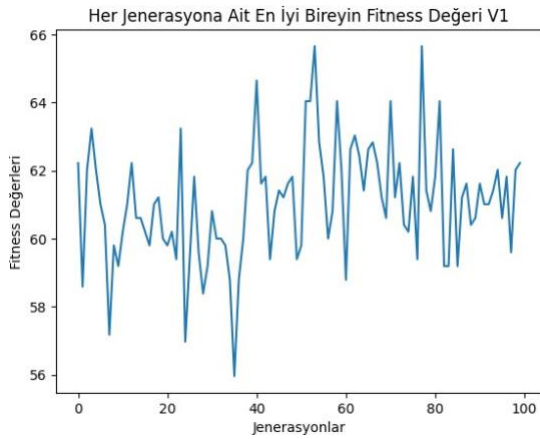
Default olan değerler:

- Popülasyondaki birey sayısı: 100
- Birey uzunluğu: 50
- Mutasyona uğrayacak gen oranı: 0.5

#### a) Mutasyona uğrayacak birey sayısı oranı: %25



#### b) Mutasyona uğrayacak birey sayısı oranı: %75



En iyi birey grafiklerine baktığımızda düşük mutasyon oranı olduğunda grafik değerlerinin diğerine göre düzgün denebilecek şekilde arttığını ve genel olarak fitness skorlarının daha yüksek olduğunu gözlemlemekteyiz. Buna sebep olarak, mutasyon oranını yüksek tuttuğumuzda elimizdeki iyi bireyleri de kaybettiğimizi söyleyebiliriz. Ortalama fitness değerlerine baktığımızda ise hem başlangıç bitiş hem de minimum maksimum değerler gözetildiğinde kayda değer değişiklik gözlemlemedik.



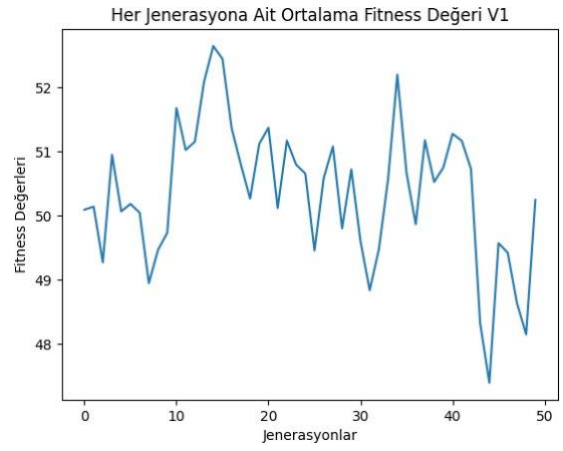
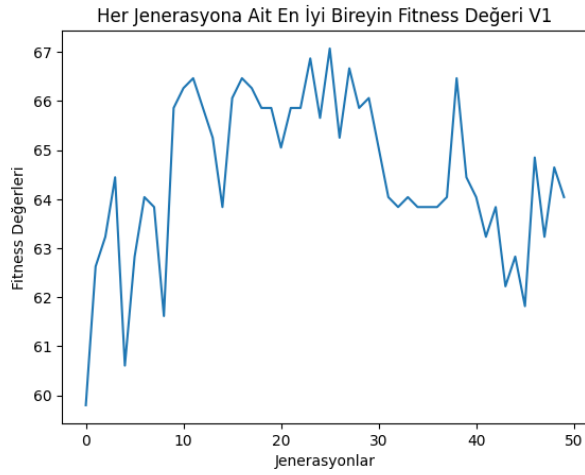
#### 4) İterasyon sayısının etkisi (Kaç jenerasyon ilerletildiği)

Default olan değerler:

- Popülasyondaki birey sayısı: 100
- Birey uzunluğu: 50
- Mutasyona uğrayacak gen oranı: 0.5
- Mutasyona uğrayacak birey oranı: 0.4

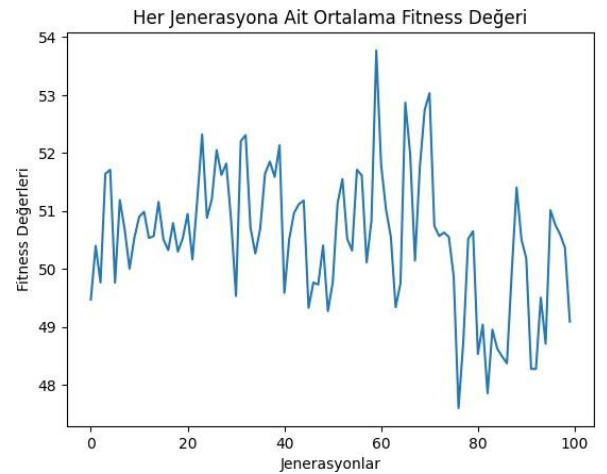
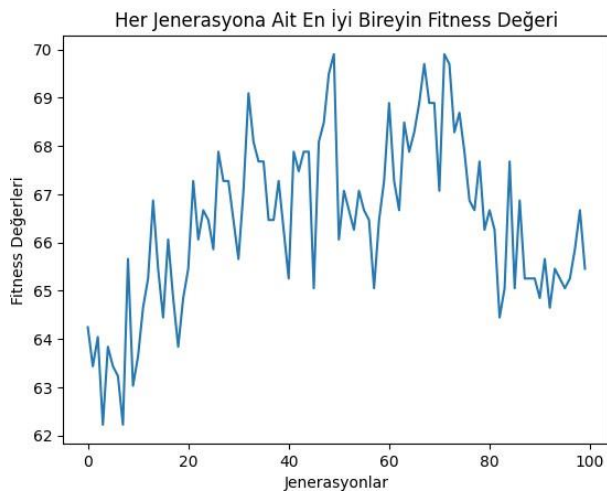
##### a) Jenerasyon Sayısı 50

(CPU 2.6 GHz iken i7 işlemci ile yaklaşık 50 sn civarlarında çalışmaktadır.)



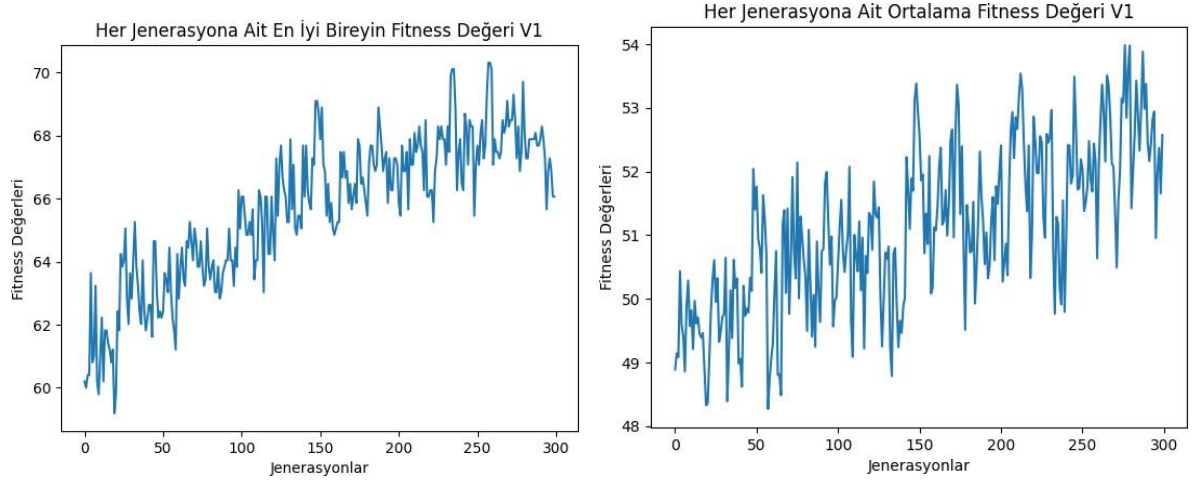
##### b) Jenerasyon Sayısı 100

(CPU 2.6 GHz iken i7 işlemci ile yaklaşık 2dk civarlarında çalışmaktadır.)



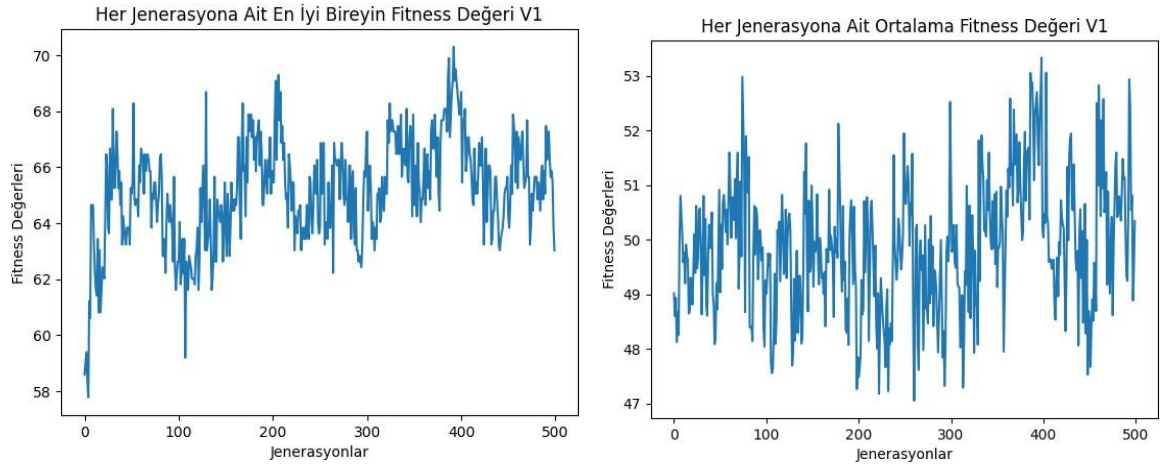
### c) Jenerasyon Sayısı 300

(CPU 2.6 GHz iken i7 işlemci ile yaklaşık 4dk civarlarında çalışmaktadır.)



### d) Jenerasyon Sayısı:500

(CPU 2.6 GHz iken i7 işlemci ile yaklaşık 9dk civarlarında çalışmaktadır.)



Jenerasyon sayılarının etkisi incelendiğinde başlangıçta iterasyon sayısını artırmanın en iyi birey grafiğinde fitness score'a olumlu etkisi olsa da bir süre sonra kayda değer bir değişiklik olmadığı, fitness score'larının benzer mertebelerde seyrettiği gözlemlenmiştir. Fonksiyonun çalışma süresini de göz önünde bulundurduğumuzda iterasyon sayısını çok yüksek tutmak manasız olacaktır.

### 3) Bulgulara Dair Genel Yorumlarımız

Genetik Arama Algoritmasında kelimeler cümledeki konumlarından ve birlikte oldukları kelimelerden bağımsız bir şekilde ele alınır. Ancak çoğu kelime bulunduğu cümle ile anlam kazanır. Bu bağlamda çalışmamıza başlarken %90 mertebelerinde başarı yüzdesine ulaşamayacağımızı tahmin etmekteydik. Nitekim genel olarak sonuçlarımızı incelediğimizde fitness score'larımızın %50-%60 bandında olduğunu, çeşitli yöntemlerle %70-%80 bantlarına çıkılabildiğini gözlemledik.

Algoritmayı daha da iyileştirmek adına, yeni jenerasyonu sadece bir önceki jenerasyonun bireyleri arasından değil, kendisinden bir ve iki önceki jenerasyonların en iyi bireylerinin belirli yüzdelerle alınmasıyla oluşturduğumuz versiyonda grafiklerimizde gözle görülür bir değişiklik gözlemleyebildik. (V2 olarak nitelendirdiğimiz versiyon) Hatta deprem veri seti üzerinde, en iyi bireyler grafiğinde fitness score'unda %85 mertebelerine çıkabildik.

Her ne kadar cross over ve mutasyon işlemleri sebebiyle her grafikte benzer iyileşmeyi elde edemesek de ata bireylerin en iyilerinin tutulması ve sonraki jenerasyonların üretiminde kullanılmasının başarı oranını artırmada kullanılabilecek önemli bir etmen olduğunu düşünmekteyiz.