



Bursa Teknik Üniversitesi

Kural Tabanlı Algoritmalar ile Duygu Analizi

Bilgisayar Mühendisliği
Mühendislik ve Doğa Bilimleri Fakültesi

Dr. Mustafa Özgür CİNGİZ
Arş. Gör. Talha KORUK

Güz 24'

İçerik

- O1 Özet
- O2 Giriş
- O3 Tasarım ve Yaklaşım
- O4 Model ve Kurallar
- O5 Uygulama ve Akış
- O6 Sonuçlar ve Testler

Proje Ekibi

Enes Şevki Dönmez: Algoritma Tasarımı/Veri Hazırlama
Kasım Deliacı: Rapor/Sunum ve Görselleştirme
Lütfü Bedel: Kodlama
Yasin Ekici: Algoritma Tasarımı/Veri Hazırlama
Yusuf Güney: Kodlama



Özet

Bu çalışmada Türkçe metinlerin duyu analizinde, dilin eklemeli yapısı ve bağlama dayalı anlam değişikliklerini göz önünde bulundurarak, daha doğru ve etkili bir analiz yöntemi sunmayı hedefledik.

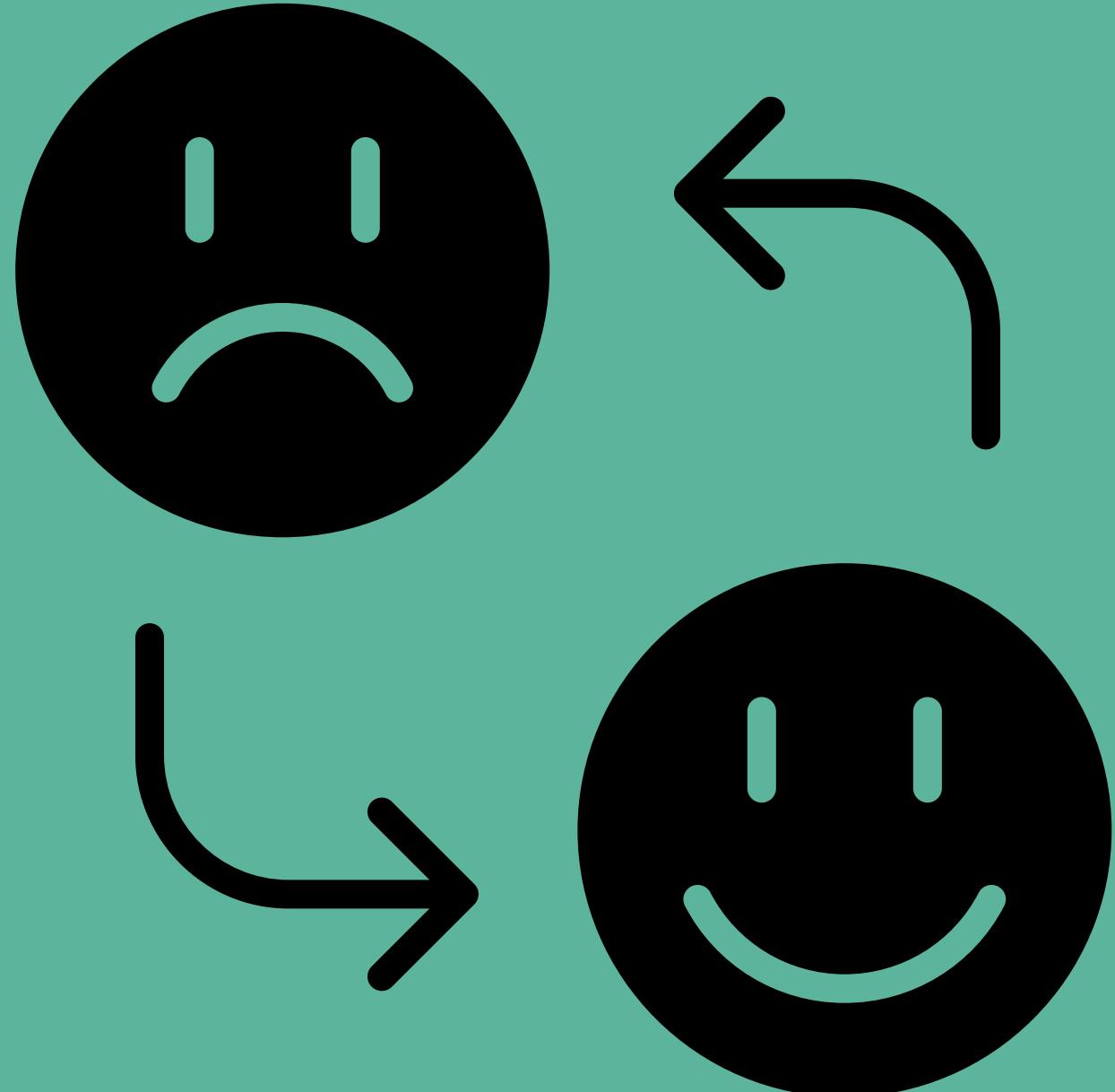
Tüm çalışmalarımız sonucunda %85'üzerinde doğruluk ve anma değerleri elde ettik.

Yöntem

- Sonlu Durum Makinesi
- Olumsuzluk Ekleri
- Çift Olumsuzluklar
- Bağlam Analizi
- Bağlaç ve Noktalamalar

Teknolojiler

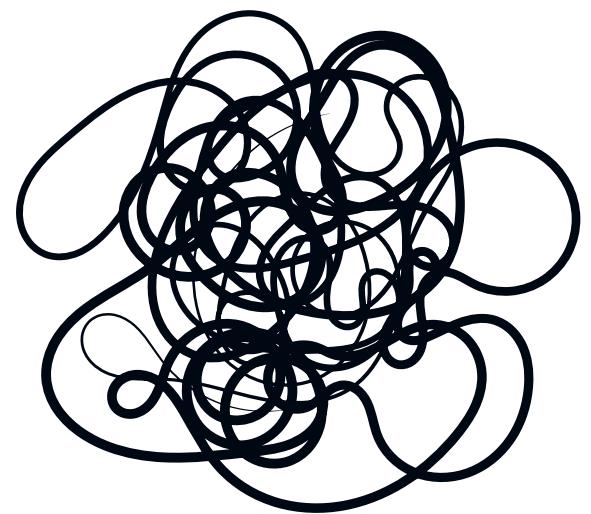
- NLTK
- Zeyrek
- Zemberek
- Regex
- OpenPyXL
- Pandas



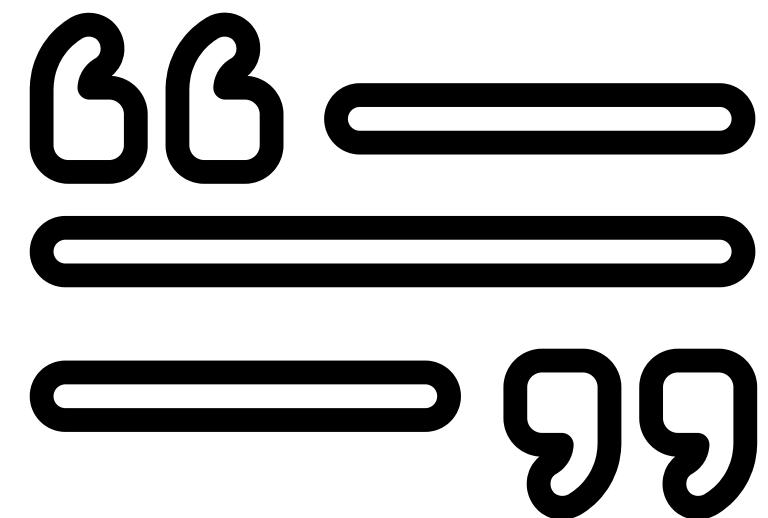
Giriş



Sondan Eklemeli Yapı



Karmaşık Bağlamlar



Karmaşık Cümle Yapısı



Tasarım?

güzel.

güzel değil.

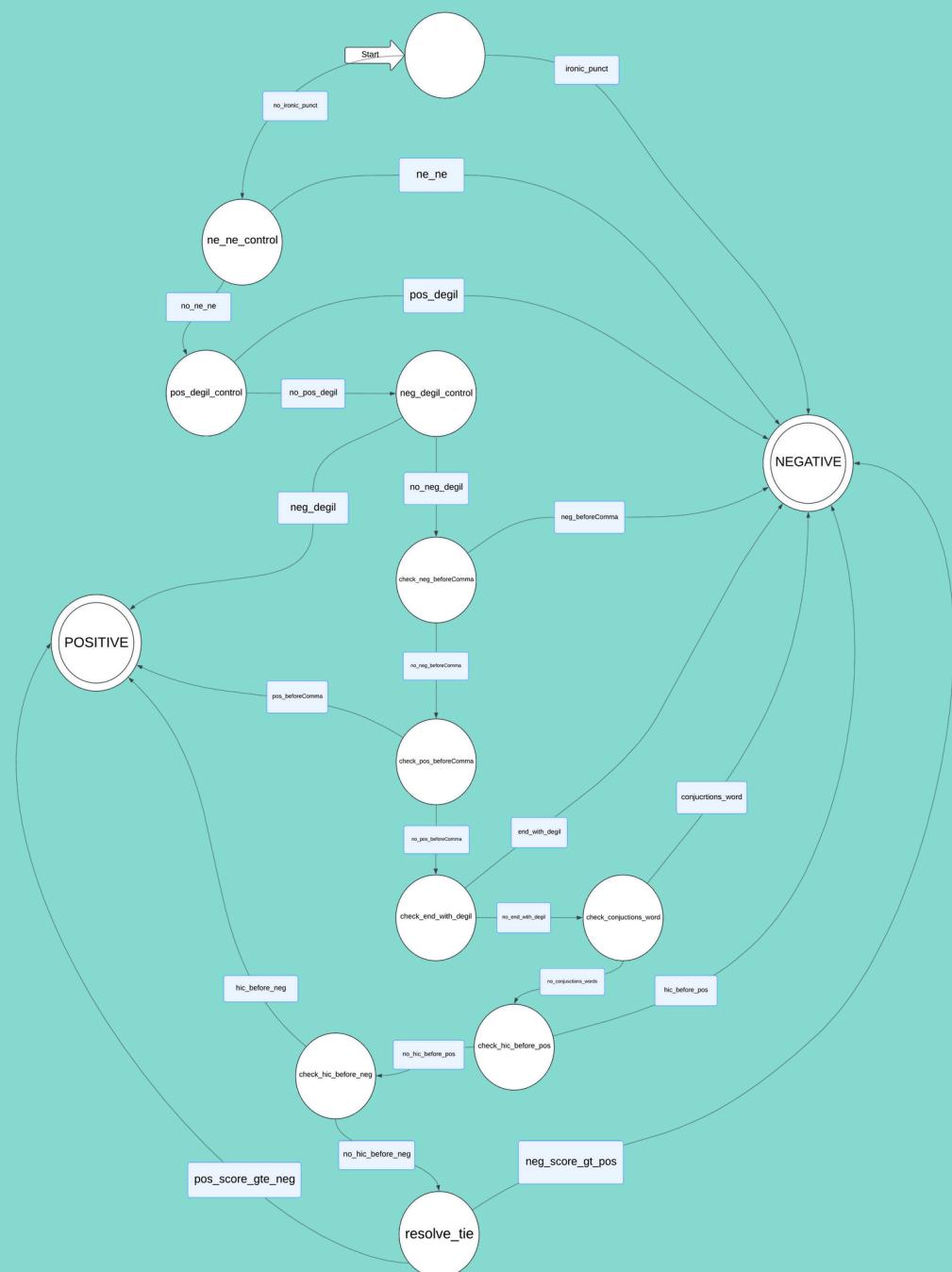
güzel değil diyemem.

Performans?

```
function register()
{
    if (!empty($_POST)) {
        $msg = '';
        if ($_POST['user_name']) {
            if ($_POST['user_password_new']) {
                if ($_POST['user_password_new'] == $_POST['user_password_repeat']) {
                    if (strlen($_POST['user_password_new']) > 5) {
                        if (strlen($_POST['user_name']) < 65 && strlen($_POST['user_name']) > 1) {
                            if (preg_match('/^[\w\d]{2,64}$/i', $_POST['user_name'])) {
                                $user = read_user($_POST['user_name']);
                                if (!isset($user['user_name'])) {
                                    if (!$_POST['user_email']) {
                                        if (strlen($_POST['user_email']) < 65) {
                                            if (filter_var($_POST['user_email'], FILTER_VALIDATE_EMAIL)) {
                                                create_user();
                                                $_SESSION['msg'] = 'You are now registered so please login';
                                                header('Location: ' . $_SERVER['PHP_SELF']);
                                                exit();
                                            } else $msg = 'You must provide a valid email address';
                                        } else $msg = 'Email must be less than 64 characters';
                                    } else $msg = 'Email cannot be empty';
                                } else $msg = 'Username already exists';
                            } else $msg = 'Username must be only a-z, A-Z, 0-9';
                        } else $msg = 'Username must be between 2 and 64 characters';
                    } else $msg = 'Password must be at least 6 characters';
                } else $msg = 'Passwords do not match';
            } else $msg = 'Empty Password';
        } else $msg = 'Empty Username';
        $_SESSION['msg'] = $msg;
    }
    return register_form();
}
```



Sonlu Durum Makinesi



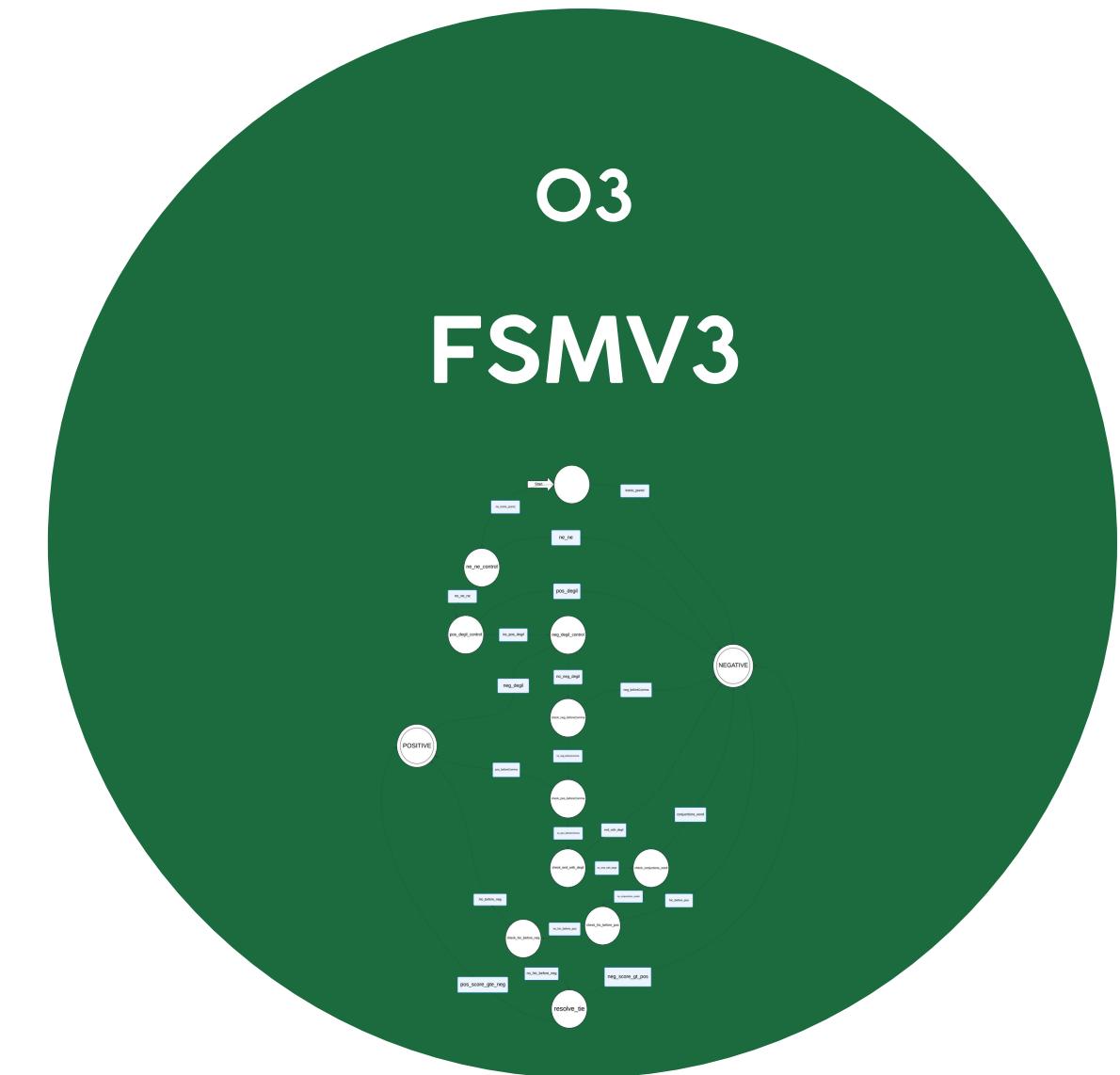
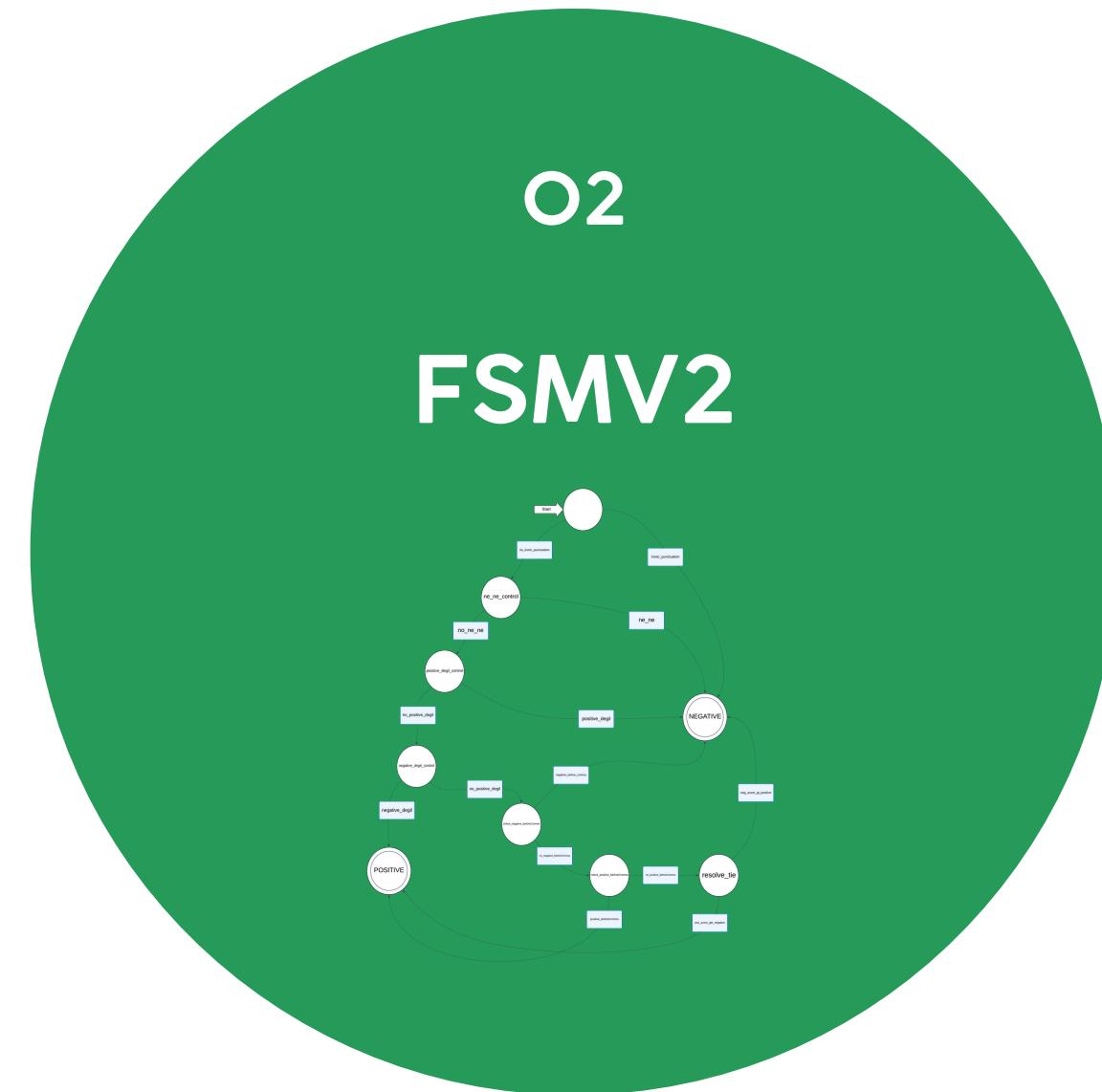
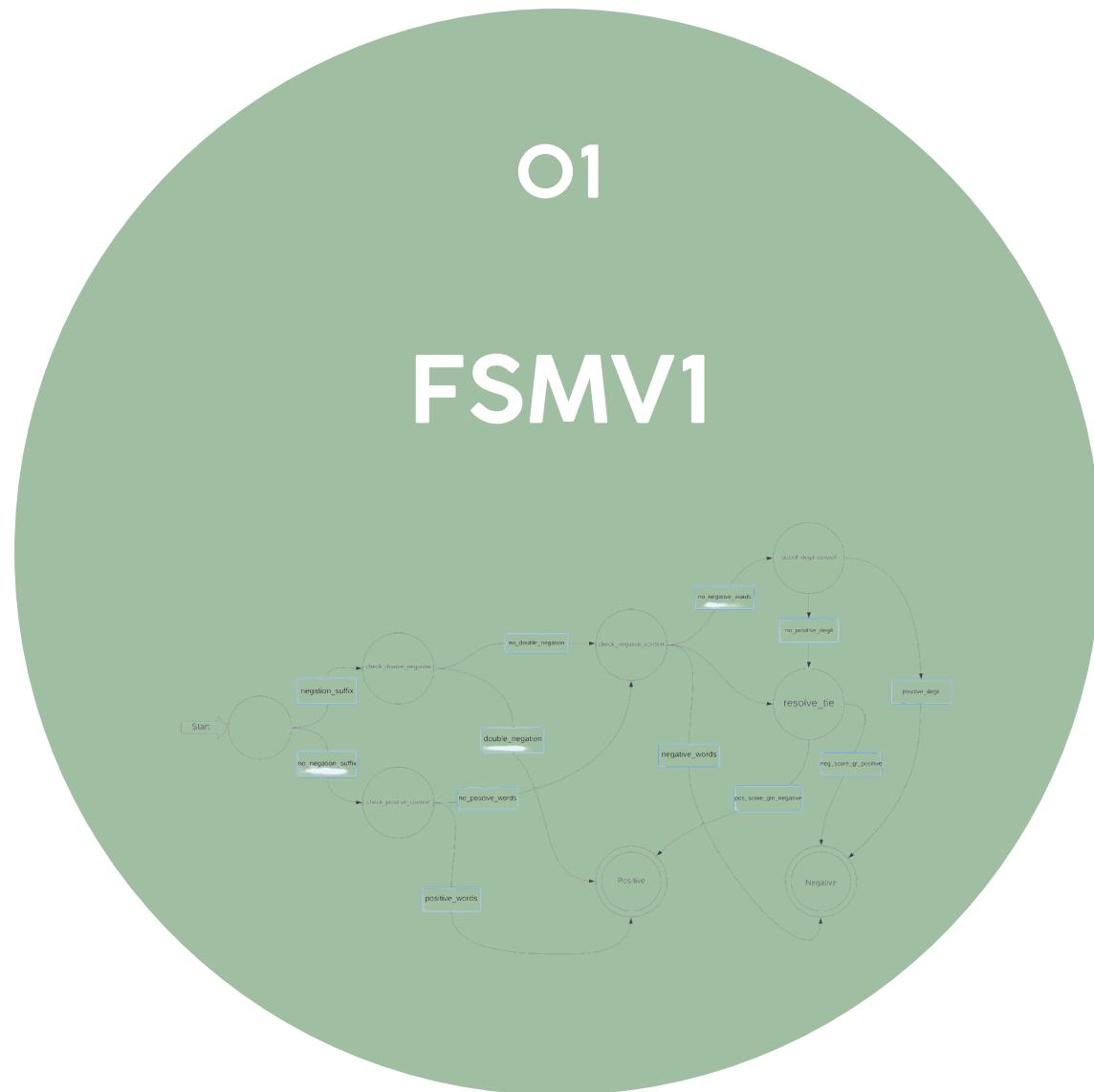
Sonlu Durum Makinesi (FSM) stratejisi, dilin karmaşık yapısını anlamlandırmak için kural tabanlı, hızlı ve açıklanabilir bir çözüm sunar.

Neden?

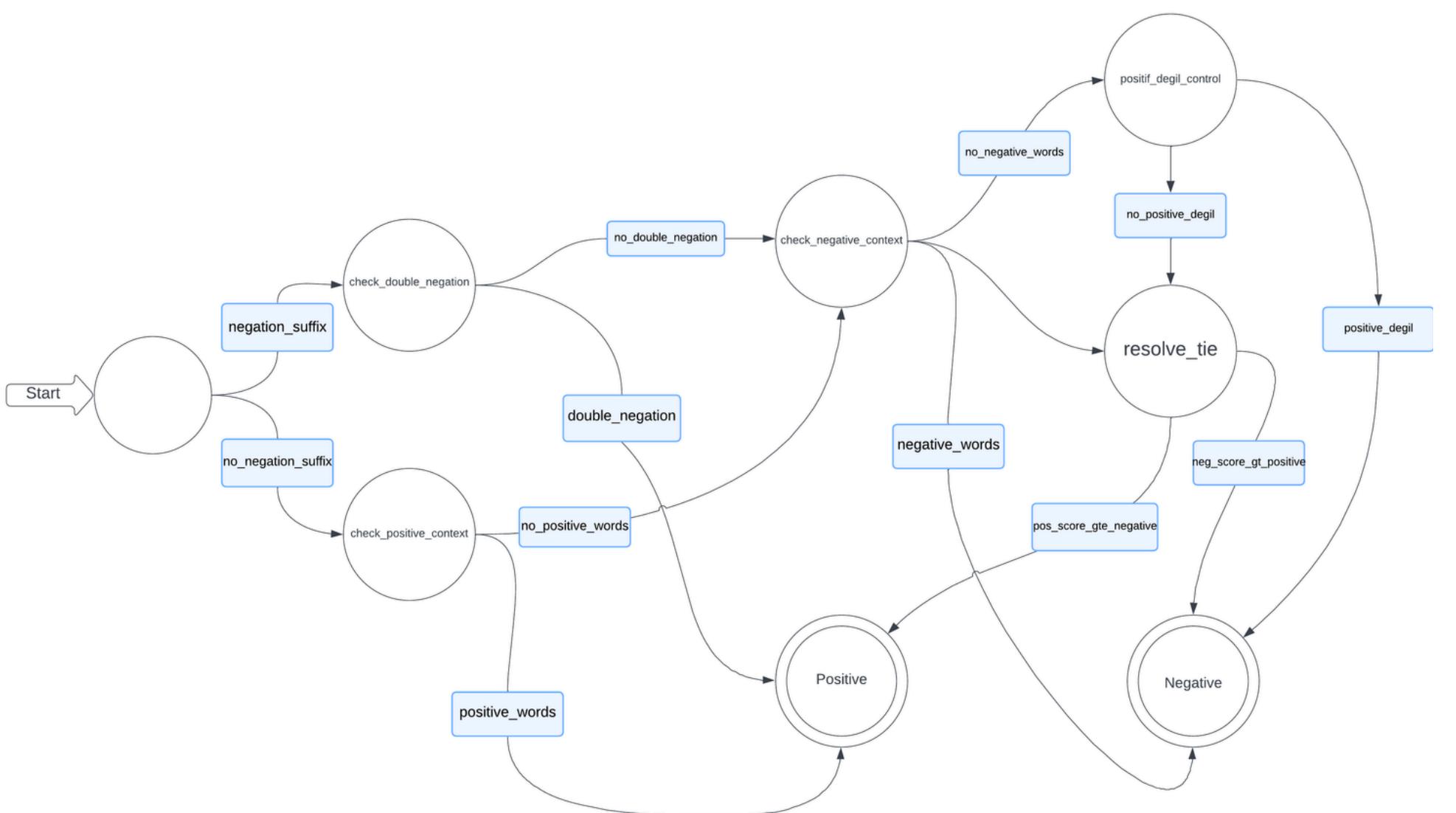
Nested if-else yapısından kaçınarak kodun okunabilirliğini ve bakımını kolaylaştırırken, performansı da artırır. Bu yöntem, metinlerin bağlamını derinlemesine analiz ederek doğru ve tutarlı sınıflandırmalar için güçlü bir temel oluşturur.



Modeller

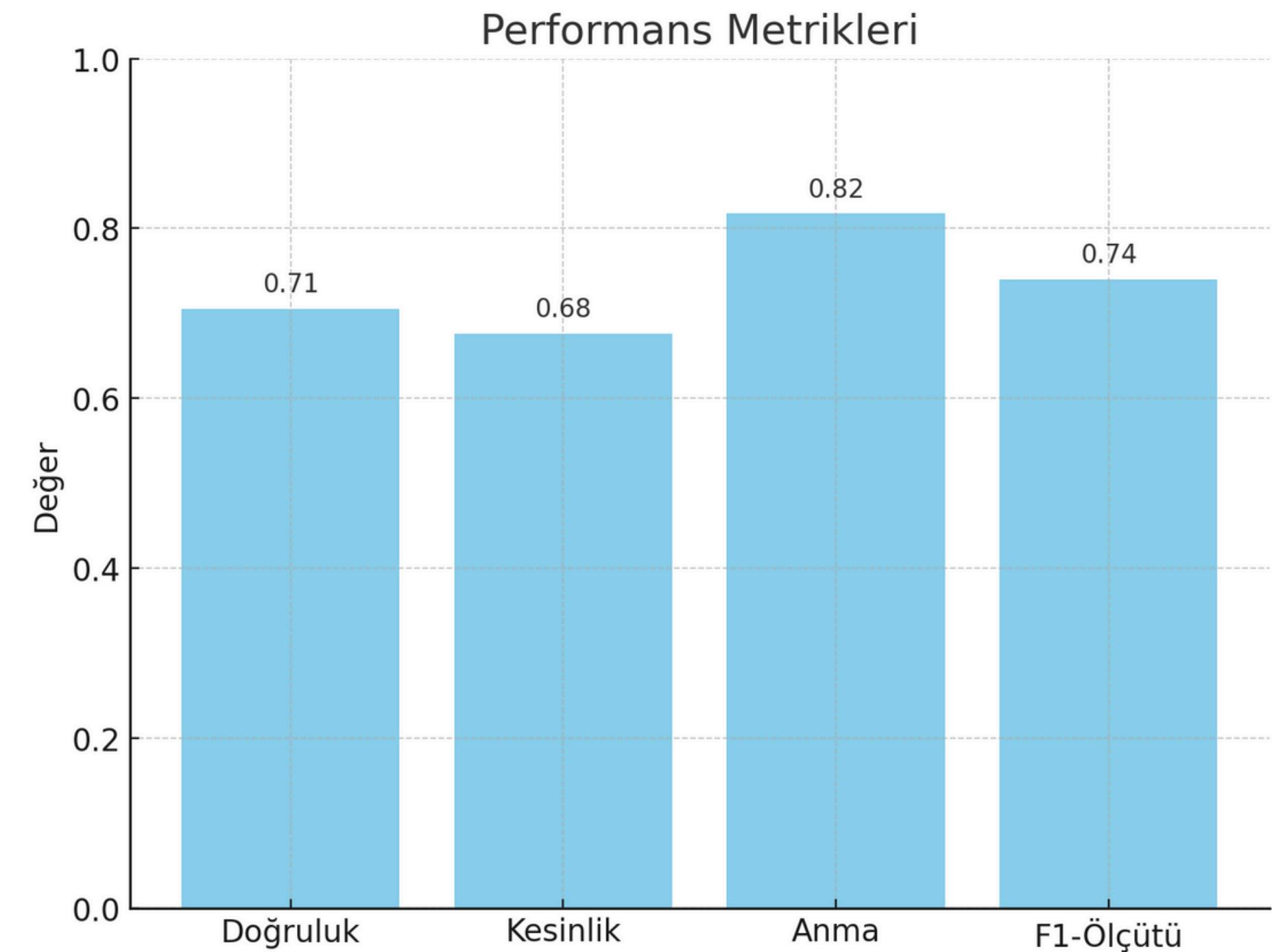
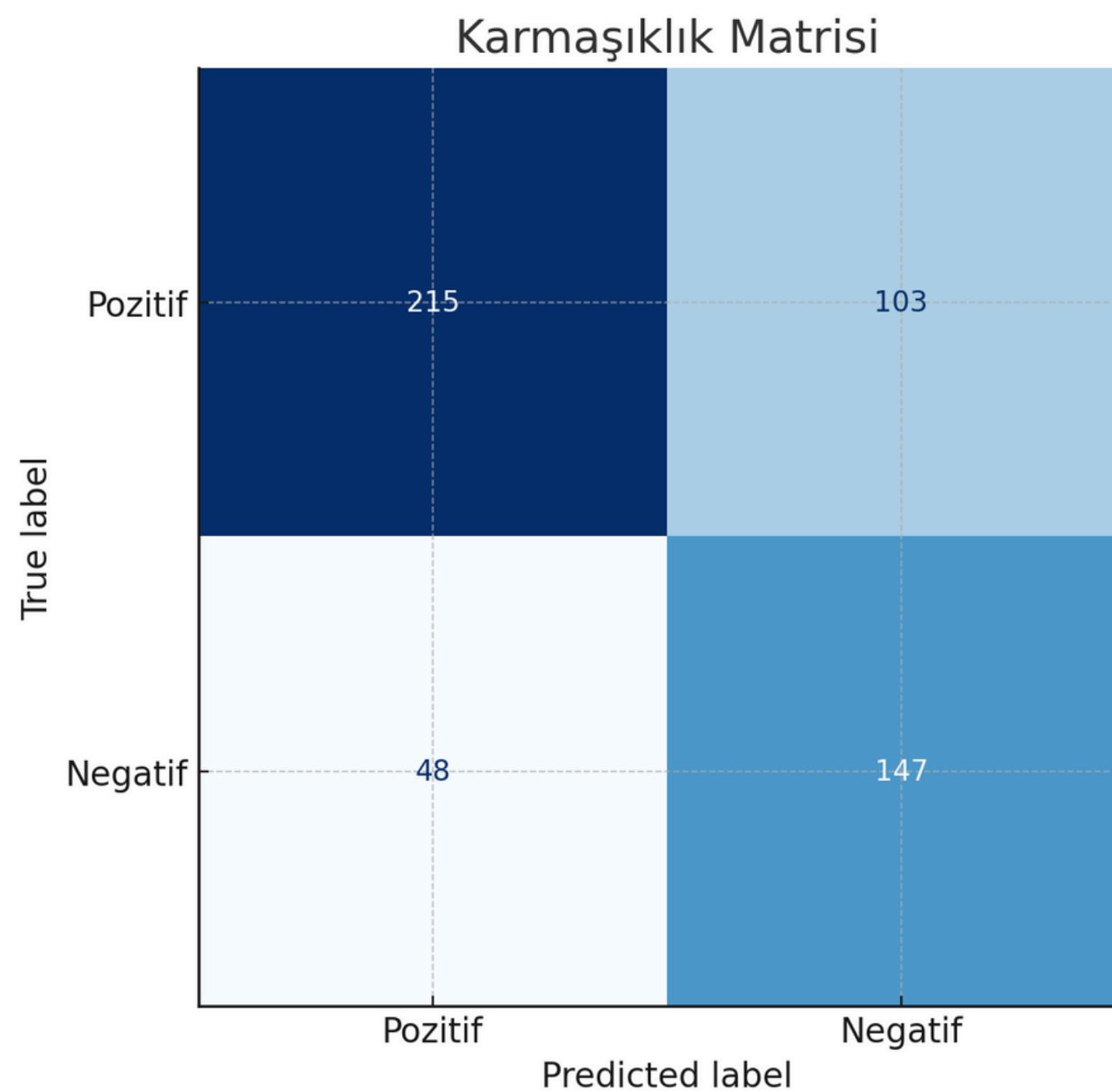


FSMV1



- 1. Start:** Başlangıç durumu, analizin ilk adımıdır.
- 2. Check Double Negation:** Çifte olumsuzlukları kontrol eder.
- 3. Check Positive Context:** Pozitif kelimelerin varlığını kontrol eder.
- 4. Check Negative Context:** Negatif kelimelerin varlığını kontrol eder.
- 5. Positive Degil Control:** Pozitif kelimelerin “değil” ile kullanılıp kullanılmadığını değerlendirir.
- 6. Resolve Tie:** Pozitif ve negatif skorları kıyaslayarak nihai sınıflandırmayı yapar.





İLK ANALİZLER

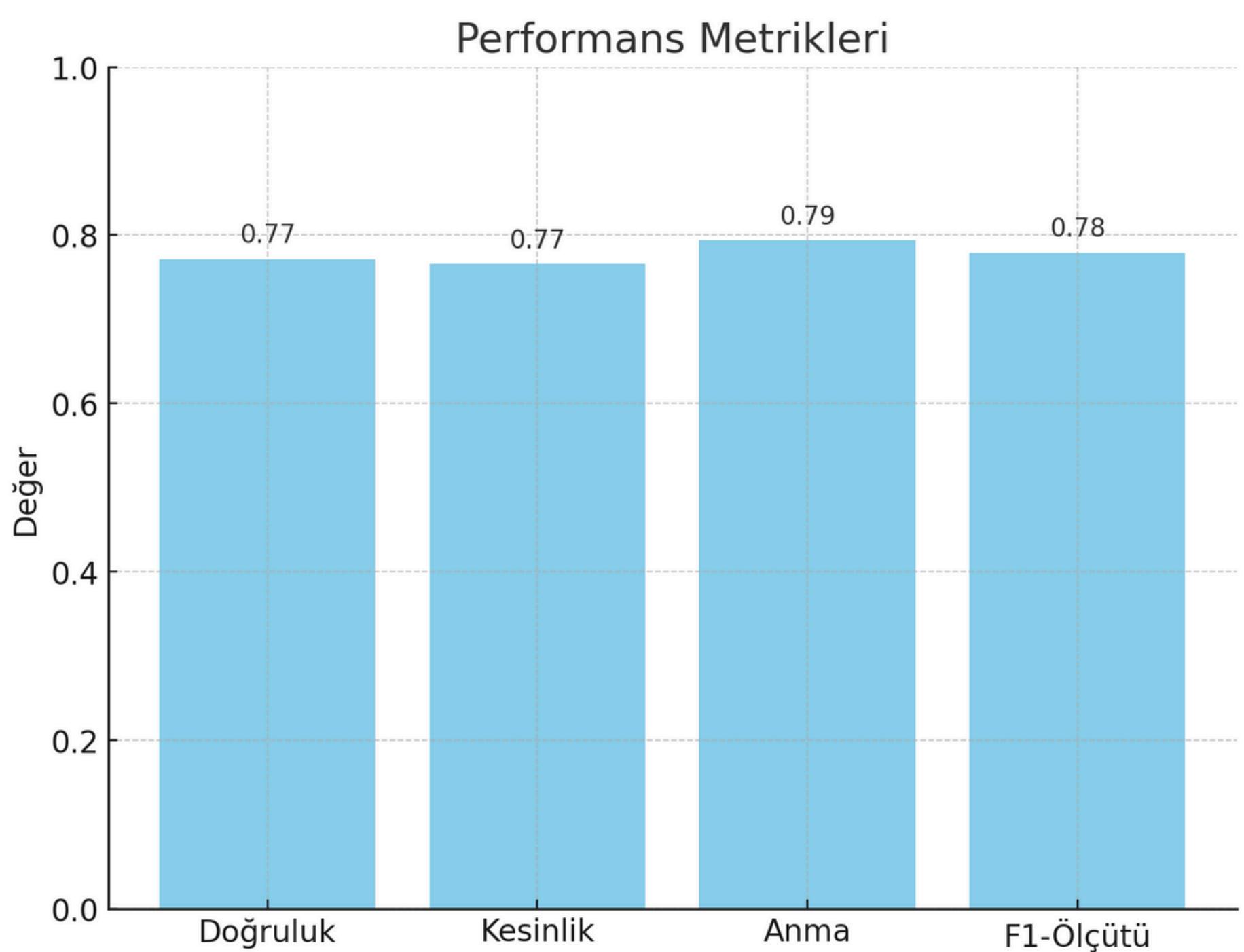
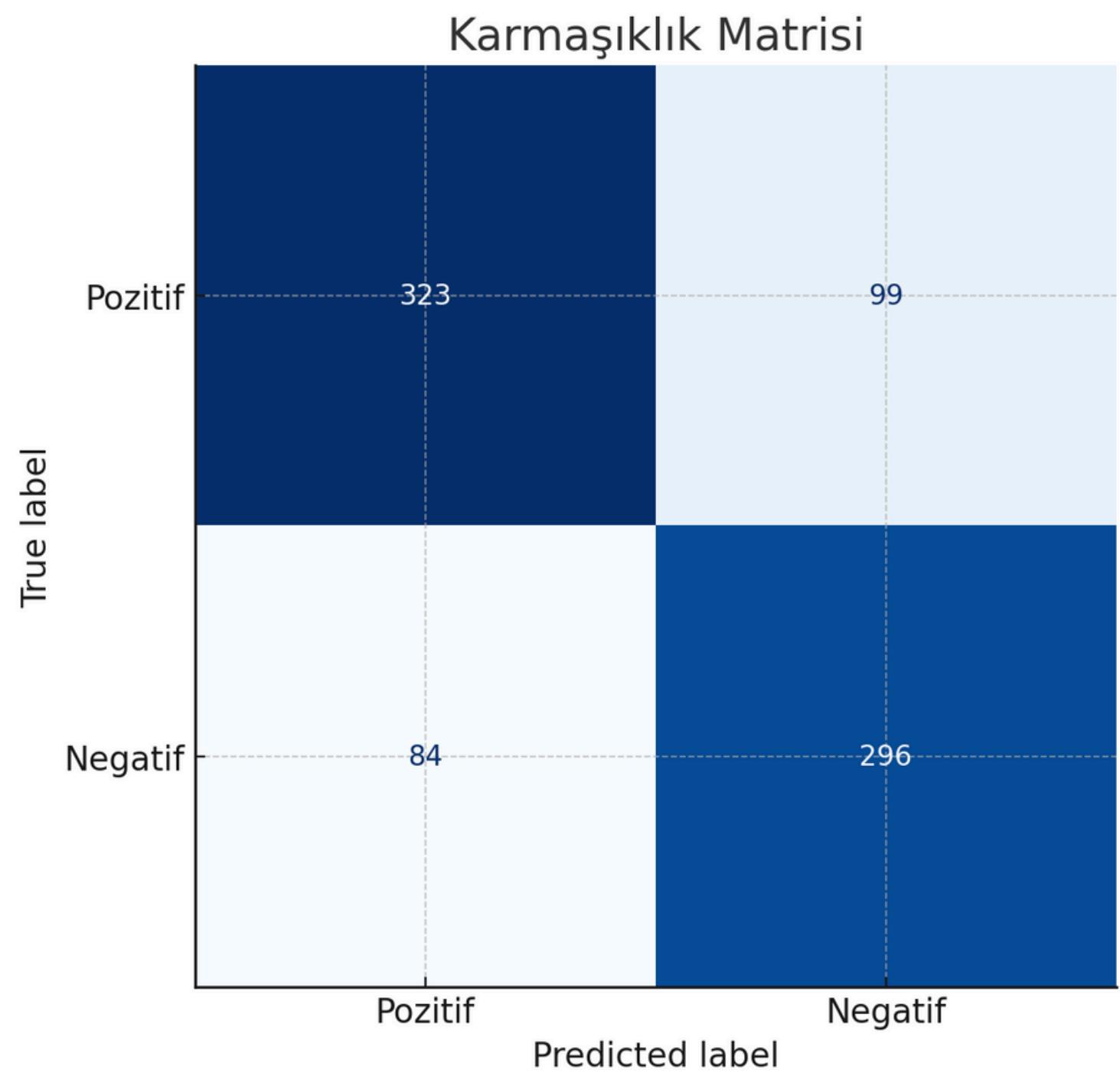
Kelime Listesinin Geliştirilmesi



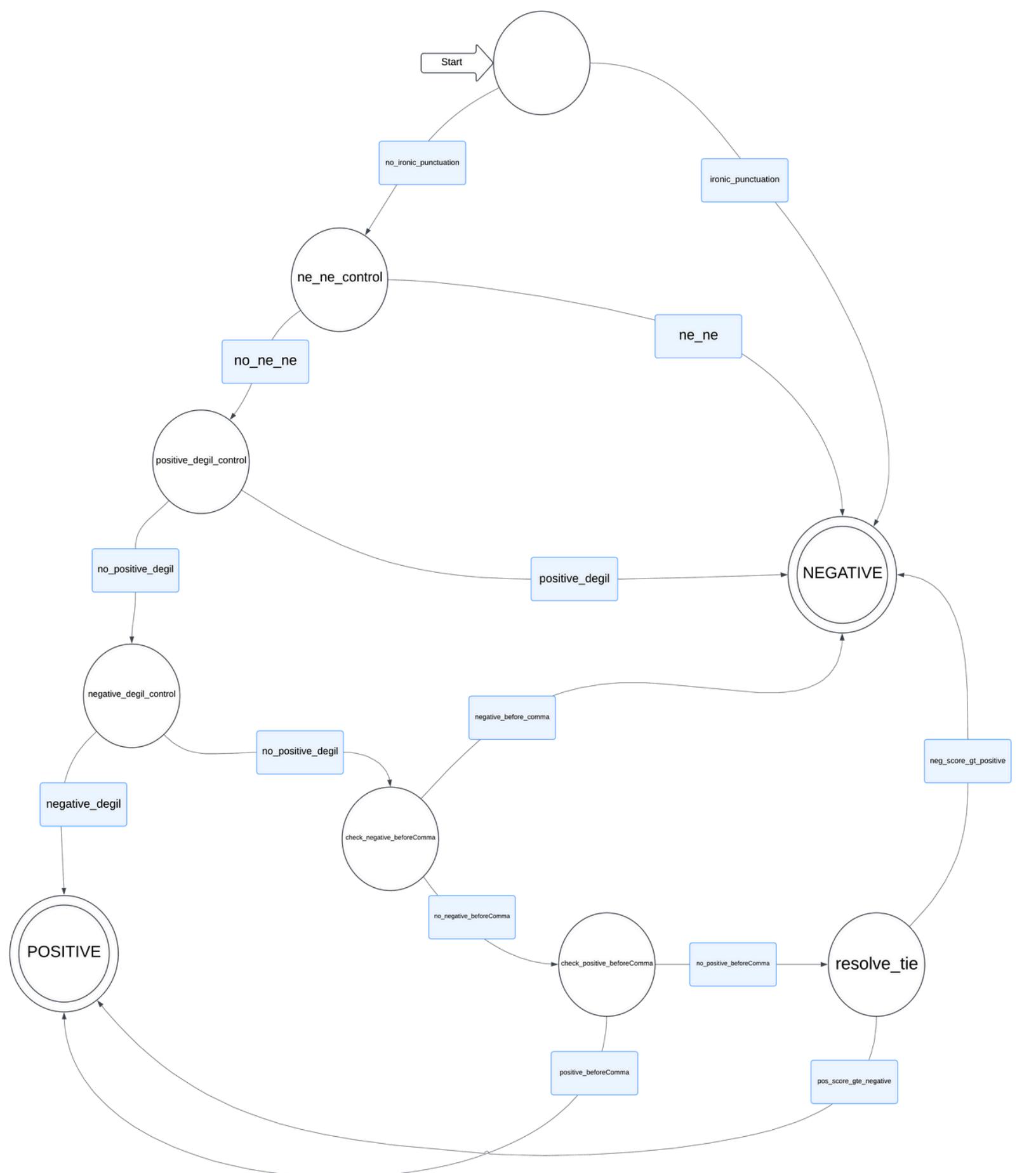
Yanlış tahminler üzerinde yapılan analizler sonucunda, FSM'nin bazı durumları doğru şekilde yakalayamadığı ve bu durumların eksik kelime listelerinden kaynaklandığı belirlenmiştir.

Pozitif ve negatif polariteli kelimelerin bulunduğu dosya, sistemin performansını artırmak amacıyla detaylı bir şekilde analiz edilerek güçlendirilmiştir.





FSMV2



1. Başlangıç Durumu (Start):
FSM analiz sürecine bu durumdan başlar ve cümlede ironi ifade eden noktalama işaretlerini kontrol eder

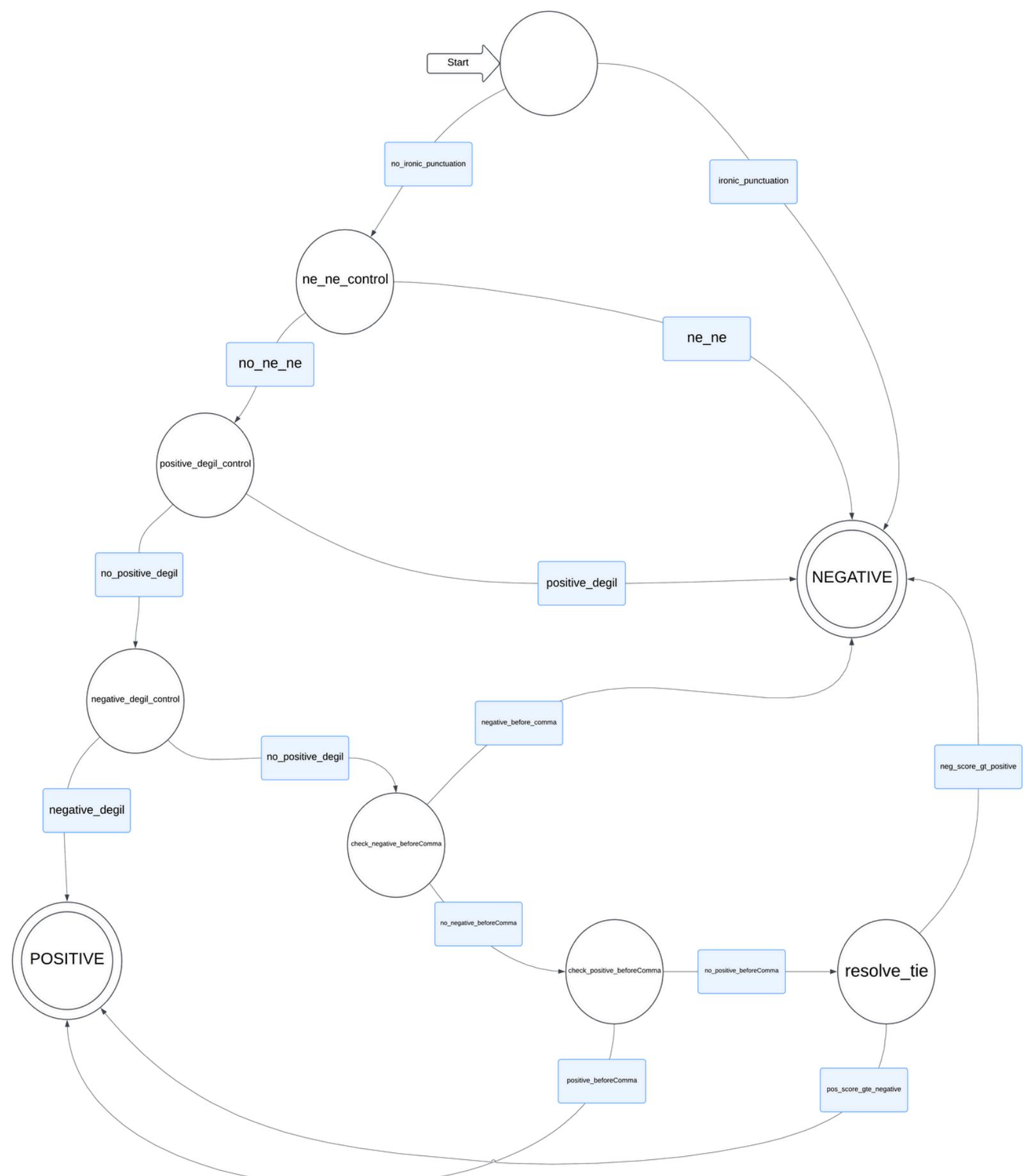
2. Ne-Ne Kontrolü (ne_ne_control):
Cümlede “ne ... ne” yapısının bulunup bulunmadığını değerlendirir.

3. Pozitif “Değil” Kontrolü (positive_degil_control):
Pozitif kelimelerin “değil” ile birlikte kullanılıp kullanılmadığını kontrol eder.

4. Negatif “Değil” Kontrolü (negative_degil_control):
Negatif kelimelerin “değil” ile birlikte kullanılıp kullanılmadığını kontrol eder.



FSMV2



5. Virgülden Önce Negatiflik Kontrolü (check_negative_beforeComma):

Virgülden önce negatif bir bağlam olup olmadığını değerlendirir.

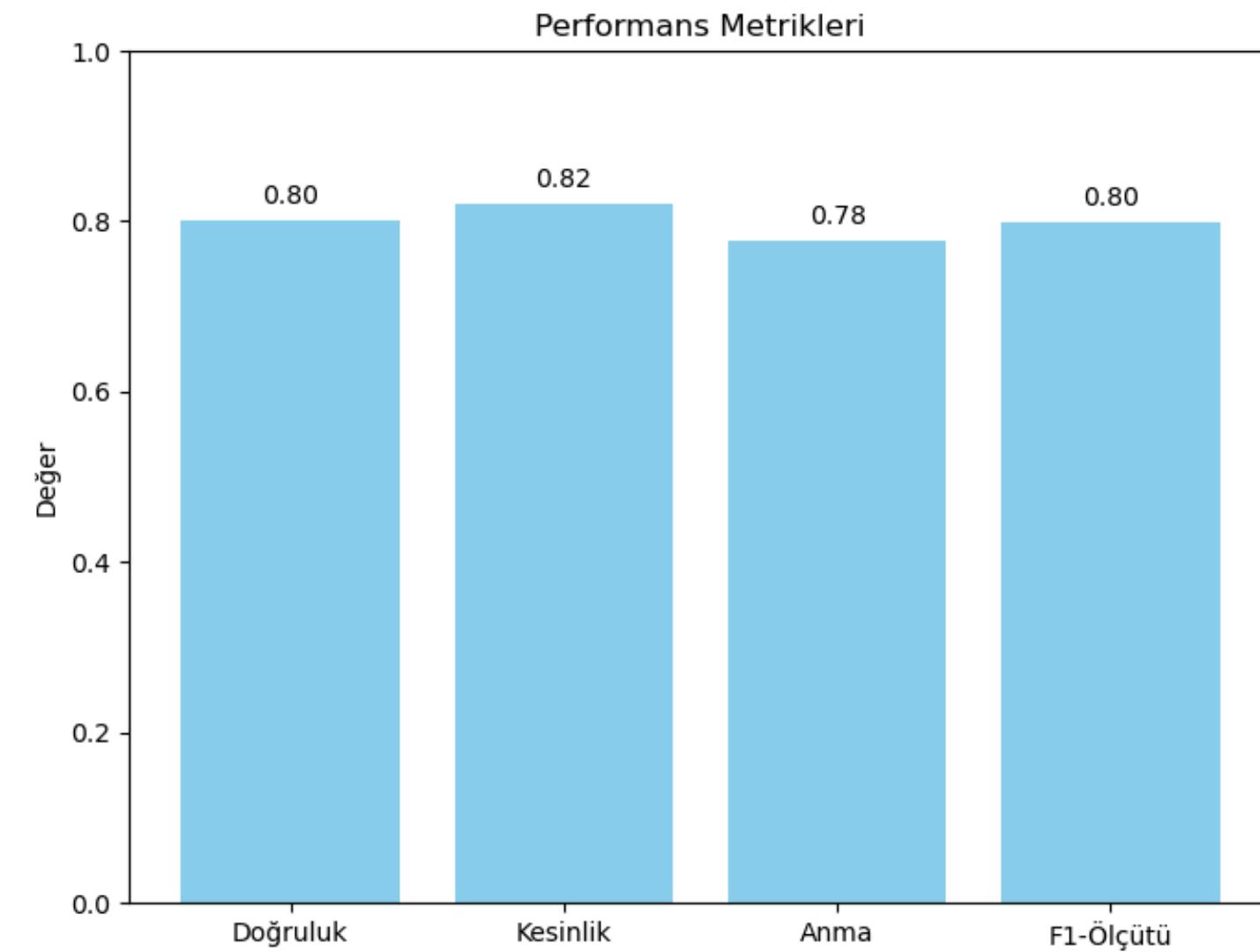
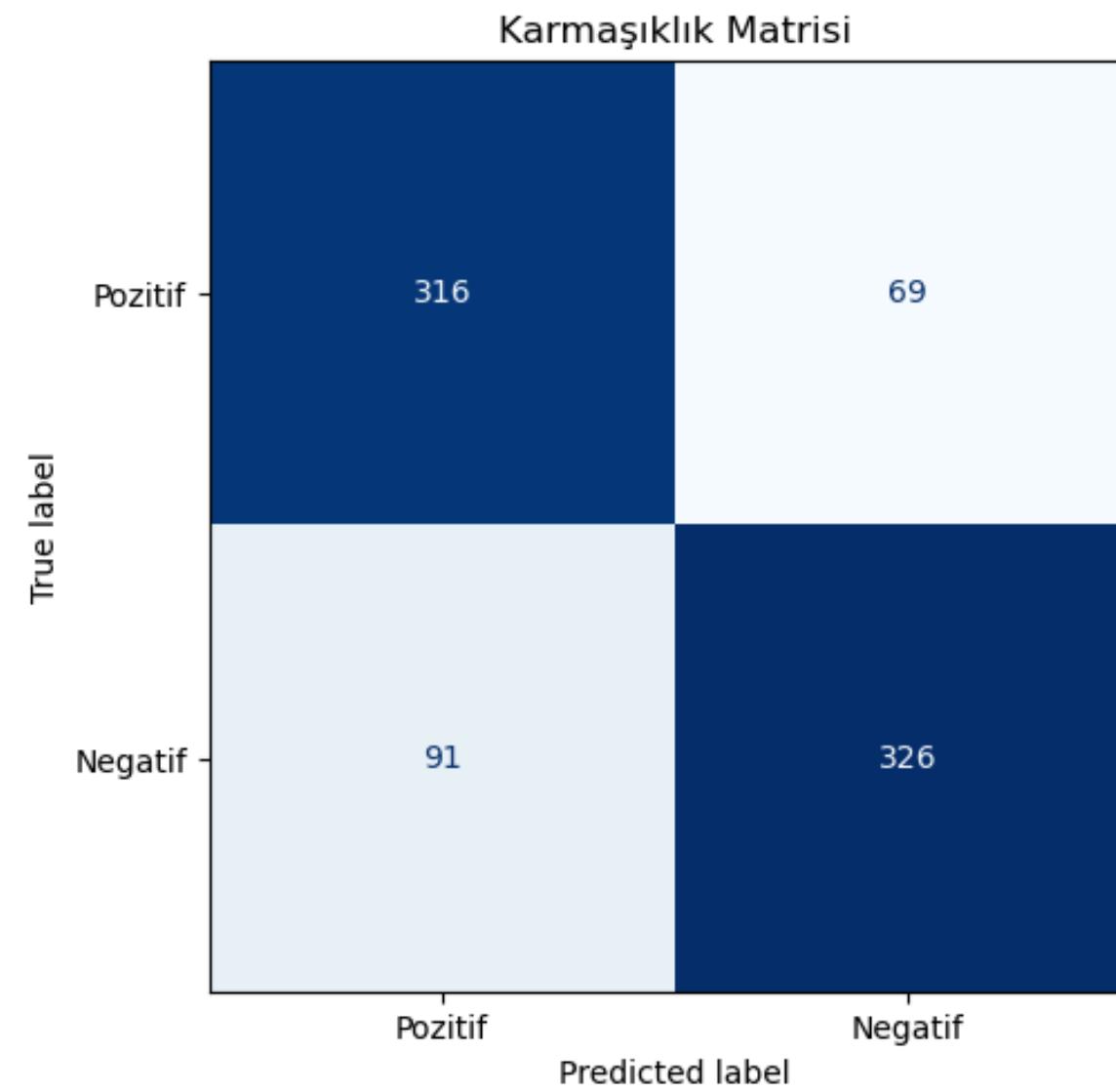
6. Virgülden Önce Pozitiflik Kontrolü (check_positive_beforeComma):

Virgülden önce pozitif bir bağlam olup olmadığını kontrol eder.

7. Skor Tabanlı Karar (resolve_tie):

Pozitif ve negatif kelime skorlarını karşılaştırarak nihai kararı verir.



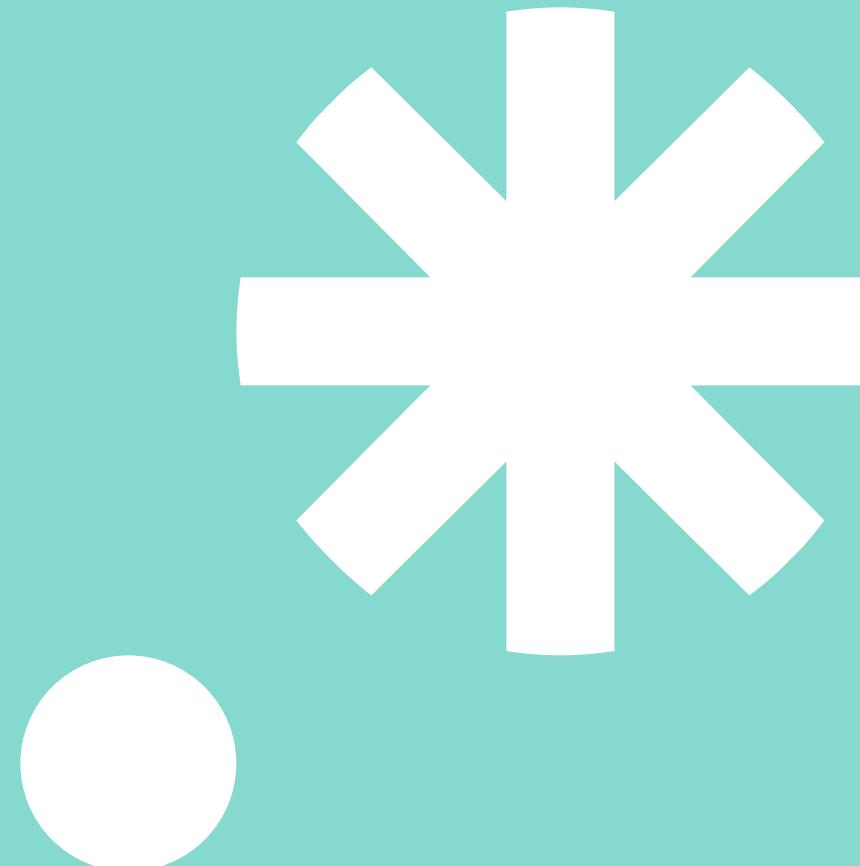


Doğruluk oranı %77'den %80'e, kesinlik %77'den %82'ye yükselerek modelin pozitif tahminlerindeki doğruluğunu artırılmıştır. Anma oranı %79'dan %78'e hafif bir düşüş göstermiş olsa da, F1-ölçütü %78'den %80'e çıkarak kesinlik ve anma dengesini daha iyi bir şekilde yansıtmaktadır.



ANALİZLER

Eklerin Ayrımı: RegEx



Hatalarımızın analizini yaparken kelimelerdeki olumsuzluk eklerini düzgün yakalayamadığımızı fark ettim.

Türkçe metinlerde olumsuzluk ekleri (-ma, -me, -mı, -mi) ile isim-fil eklerini (-mayı, -meye) doğru bir şekilde ayırt ederek performansı artırma yoluna gittik

Bu ayrımı yapabilmek adına RegEx'den faydalandık.



1. İsim-Fiil Eklerini Tespit Eden Regex:

`r'^^(.*?)(ma|me|mı|mi)([yğ][ıiüuae]).*'`

- $(ma|me|mı|mi)$ → ma, me, mı, mi eklerini hedefler.
- $[yğ][ıiüuae]$ → Bu ekten sonra gelen bir y veya ğ harfiyle başlayıp bir ünlü harfle devam eden yapıları yakalar (-mayı, -meye, -meyi).
- $.*$ → Ekin ardından gelen herhangi bir karakter dizisini kabul eder.
- $^$ ve $$$ → Kelimenin baştan sona taranmasını sağlar.

Deseni Daraltmak

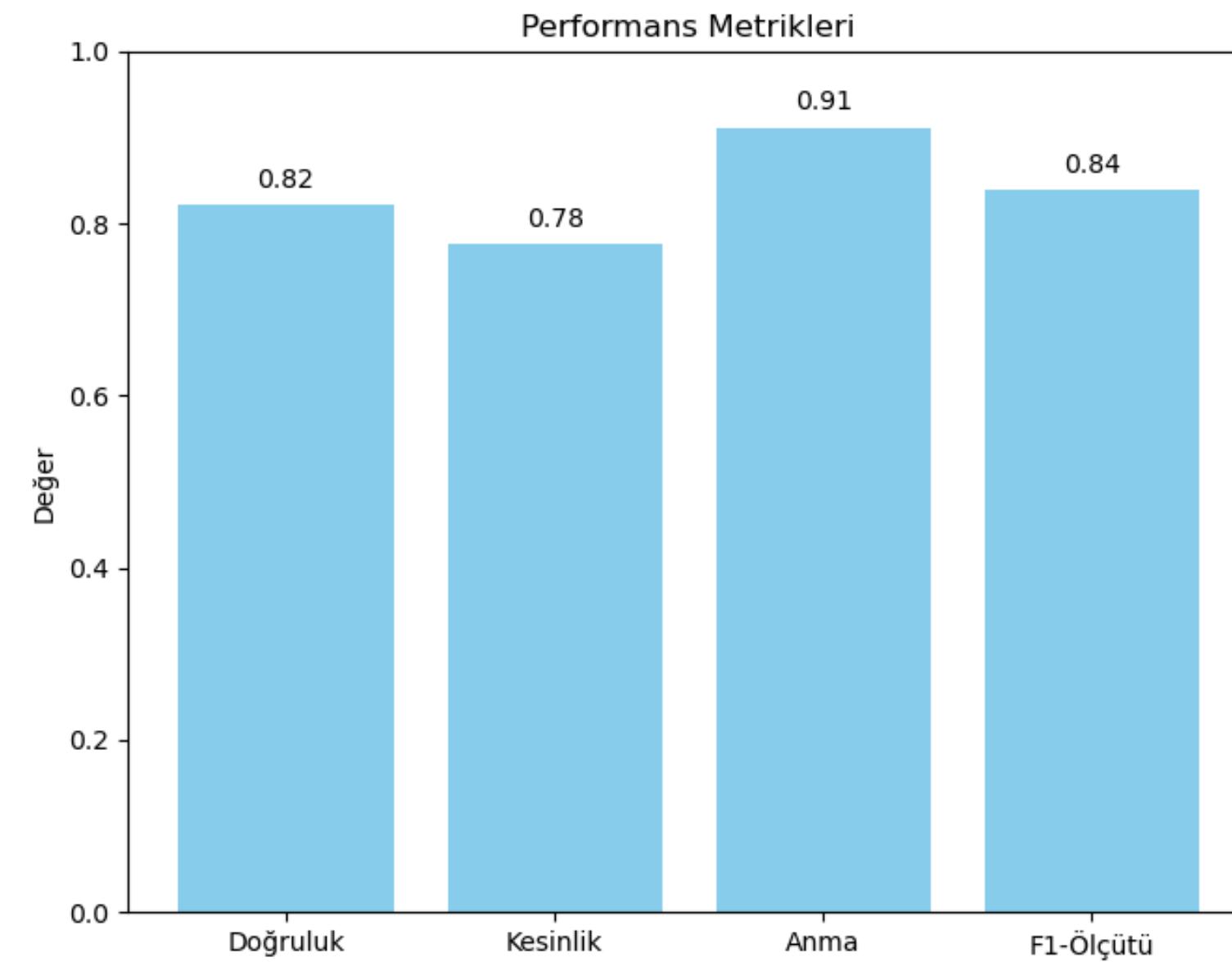
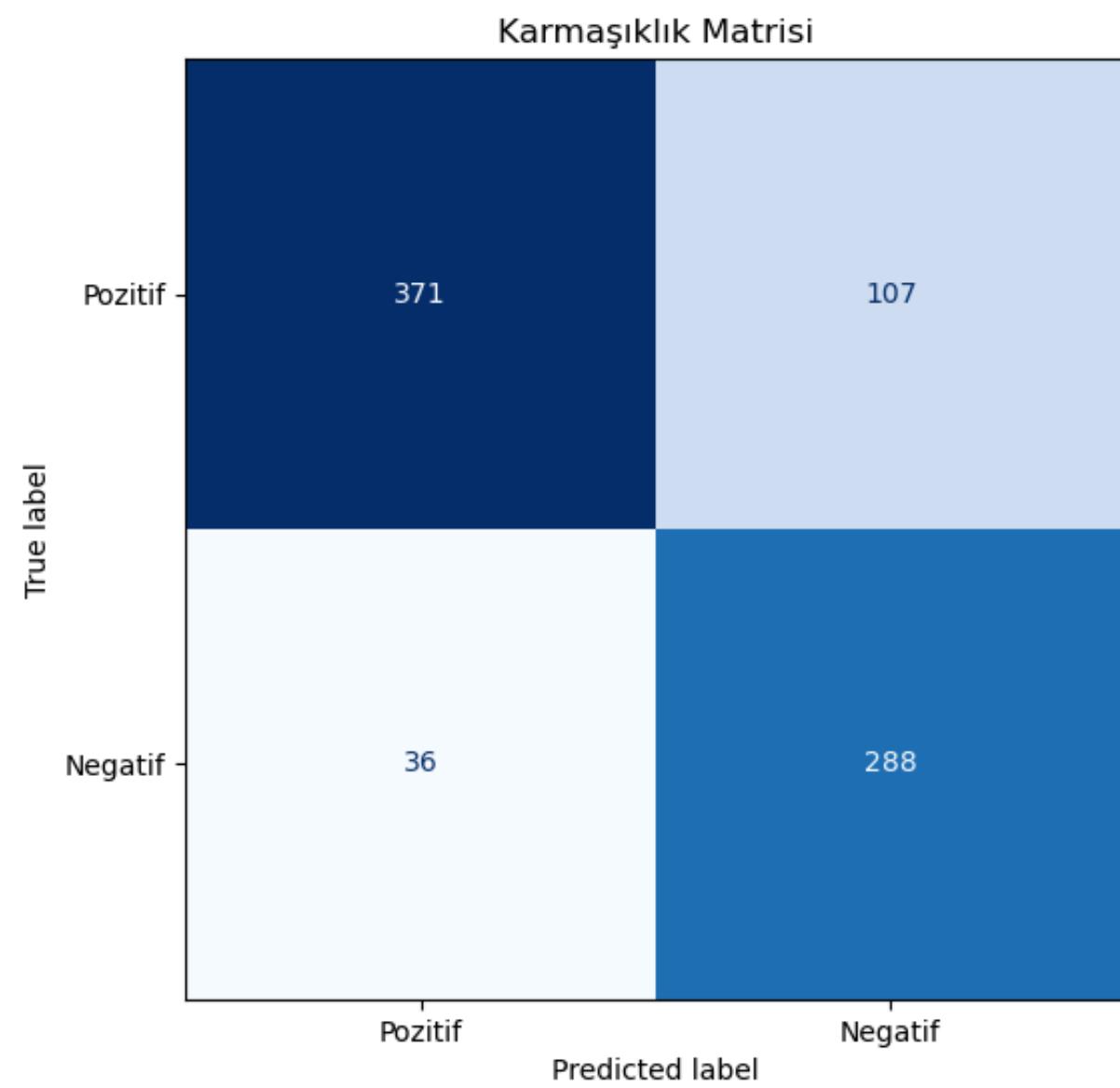
`r'^^(.*?)(ma|me)(y[ıiüuae])(?!n).*'`

- `r'^^(.*?)(ma|me)(y[ıiüuae])(?!n).*`
- $(?!n)$ → Lookahead kullanılarak, yi veya yi gibi eklerden sonra n harfi gelmesi durumunda eşleşme yapılmaz. (koşmayın)

2. Olumsuzluk Eklerini Tespit Eden Regex:

`r'^^(.*?)(ma|me|mı|mi).*'`

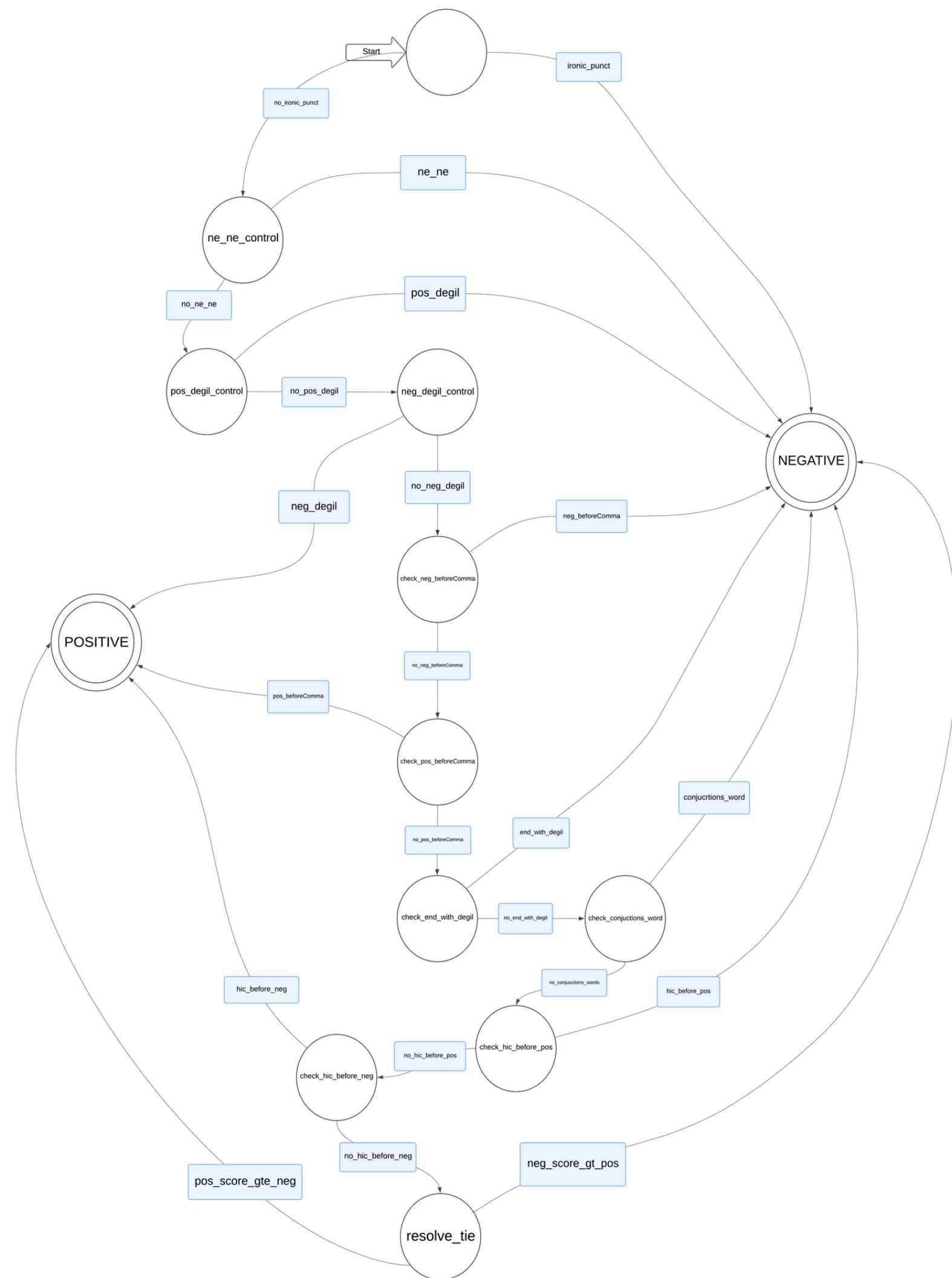
- $ma|me|mı|mi$ → Kelimenin herhangi bir yerinde ma, me, mı, mi eklerini arar.
- Bu desen, olumsuzluk anlamı katan yapıları (ör. okumadım, sevmiyor) doğru şekilde tespit eder.



Doğruluk oranı %82'ye yükselmiş ve model genel olarak daha doğru tahminler yapmıştır. Kesinlik %78'e düşmesine rağmen, anma oranındaki %91'e çıkan artış, modelin pozitif örnekleri yakalama kapasitesini önemli ölçüde artırdığını göstermektedir. F1-ölcütü ise %84'e çıkarak kesinlik ve anma arasındaki dengenin daha da iyileştiğini ifade etmektedir.

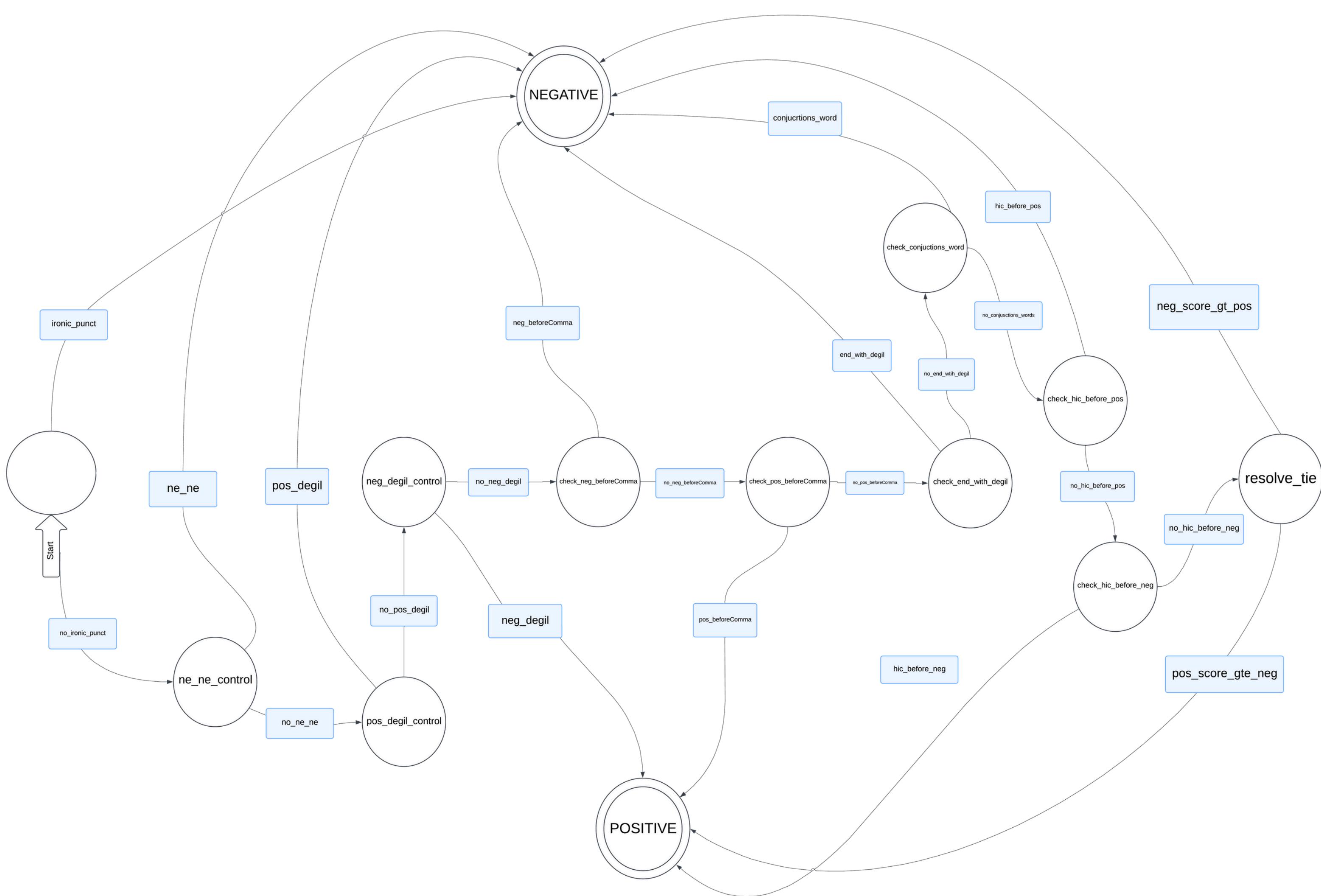


FSMV3: Final Model



ilk olarak temel bir FSM modeli geliştirildi ve performansı analiz edildi. Daha sonra, pozitif ve negatif kelime listesi güçlendirilerek modelin bağlam algısını iyileştirildi. Ardından, eksikleri giderdiğimiz yeni bir FSM modeli tanıttık, bu modelin gelişiminin üzerine regex ile ekleri daha doğru ayırtırarak duyarlılığı arttırdık şimdi eski yaklaşımın güclü yönlerini koruyarak daha geniş bağamları ve dil yapısını etkili bir şekilde ele almayı hedefleyen final FSM modelimizi tanıtacağız.





FSMV3: Final Model

1. Başlangıç Durumu (start)

FSM analizine başlangıç noktasıdır. Burada, cümlenin genellikle en üst düzeyde anlamını etkileyen özellikler kontrol edilir.

Kural:

ironic_punctuation: Cümlede ironi veya alay ifade eden bir noktalama işaretleri varsa (örneğin, (!)) FSM doğrudan "Negatif" durumuna geçer.

no_ironic_punctuation: İronik bir noktalama işaretri yoksa, FSM "ne_ne_control" durumuna geçer.

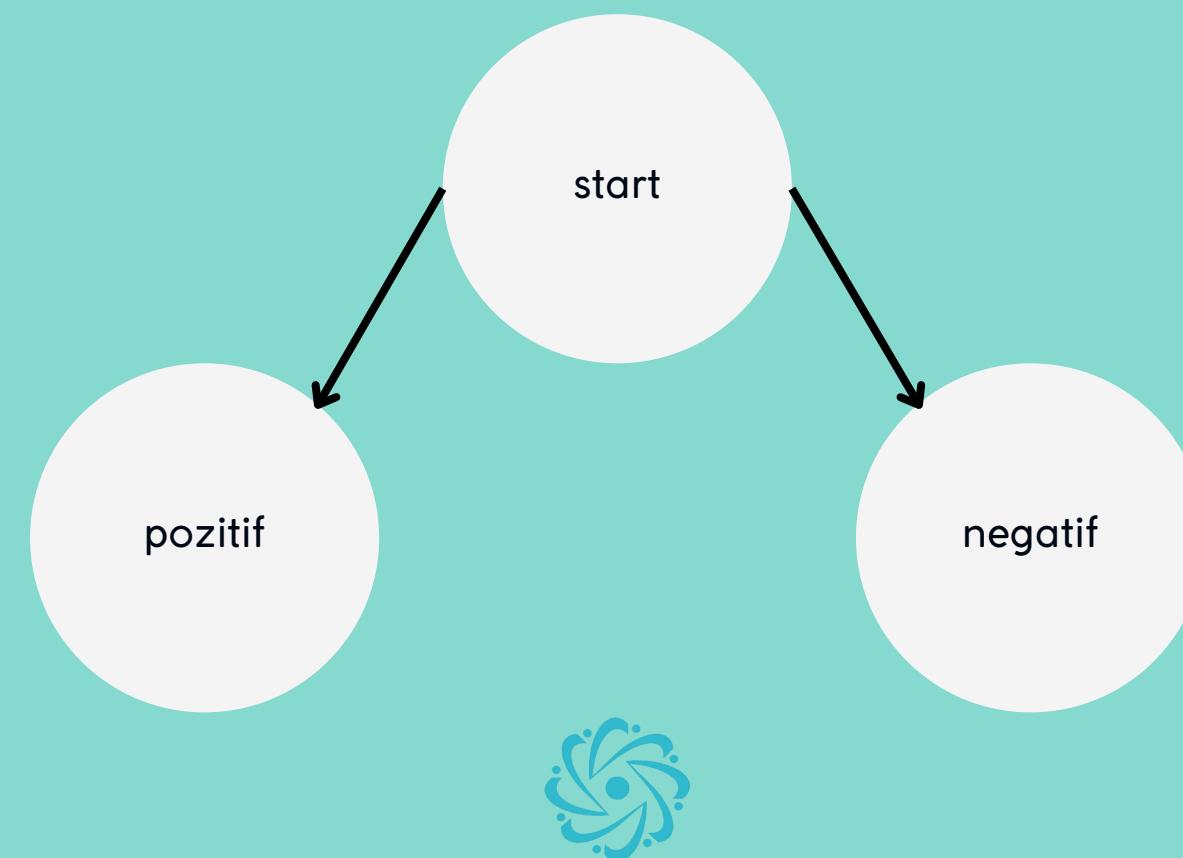
2. Ne-Ne Kontrolü (ne_ne_control)

Bu durum, Türkçe'de sıkça kullanılan "ne ... ne" yapısını değerlendirir.

Kural:

ne_ne: Eğer "ne ... ne" yapısı varsa (örneğin, "ne güzel ne kötü"), FSM doğrudan "Negatif" durumuna geçer.

no_ne_ne: Böyle bir yapı yoksa, FSM "positive_degil_control" durumuna geçer.



FSMV3: Final Model

3. Pozitif "Değil" Kontrolü (**positive_degil_control**)

Cümlede pozitif kelimelerin "değil" ile birlikte kullanılıp kullanılmadığı değerlendirilir.

Kural:

positive_degil: Pozitif kelimeler "değil" ile birlikte kullanılmışsa, FSM "Negatif" durumuna geçer.

no_positive_degil: Böyle bir kullanım yoksa, FSM "negative_degil_control" durumuna geçer.

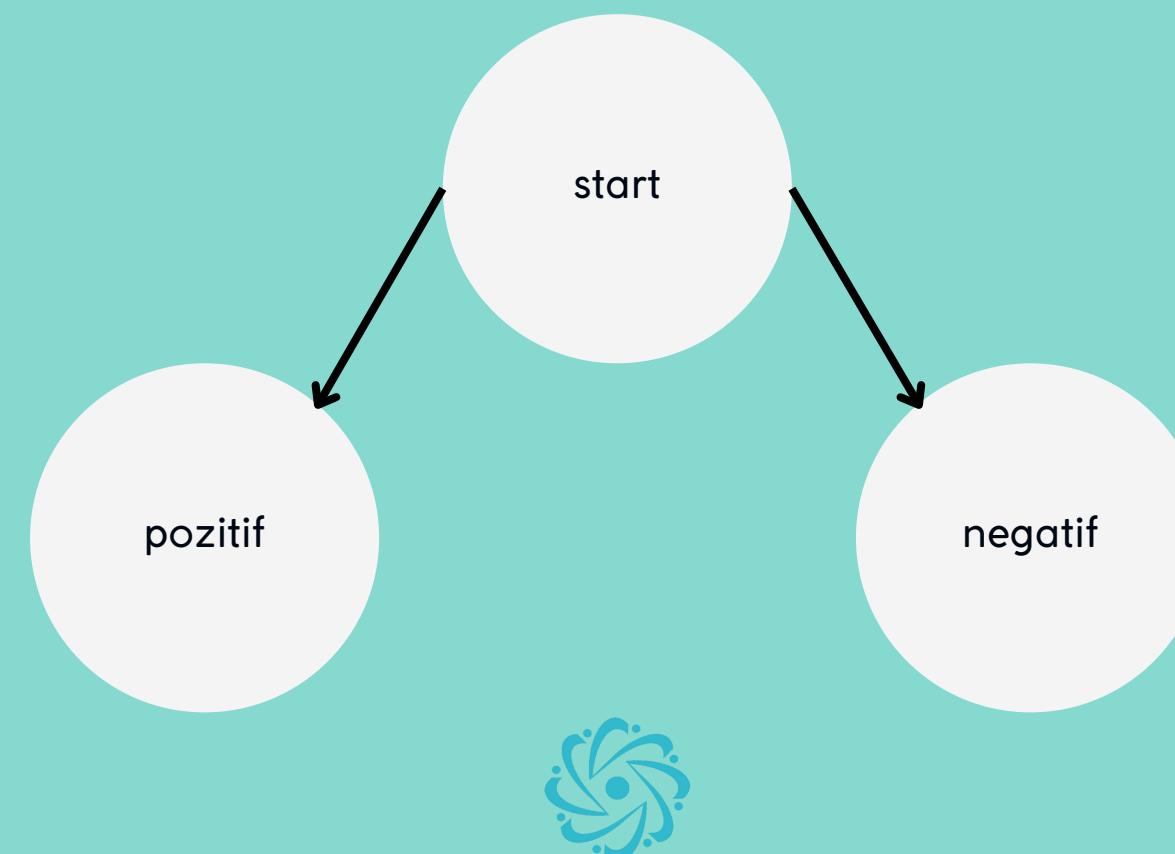
4. Negatif "Değil" Kontrolü (**negative_degil_control**)

Negatif kelimelerin "değil" ile birlikte kullanılıp kullanılmadığını değerlendirir.

Kural:

negative_degil: Negatif kelimeler "değil" ile birlikte kullanılmışsa, FSM "Pozitif" durumuna geçer.

no_negative_degil: Böyle bir kullanım yoksa, FSM "check_negative_beforeComma" durumuna geçer.



FSMV3: Final Model

5. Virgülden Önce Negatiflik Kontrolü (check_negative_beforeComma)

Cümlede bir virgülden önce negatif bir bağlamın olup olmadığı kontrol edilir.

Kural:

negative_beforeComma: Eğer negatif bir bağlam varsa, FSM "Negatif" durumuna geçer.

no_negative_beforeComma: Negatif bağlam yoksa, FSM "check_positive_beforeComma" durumuna geçer.

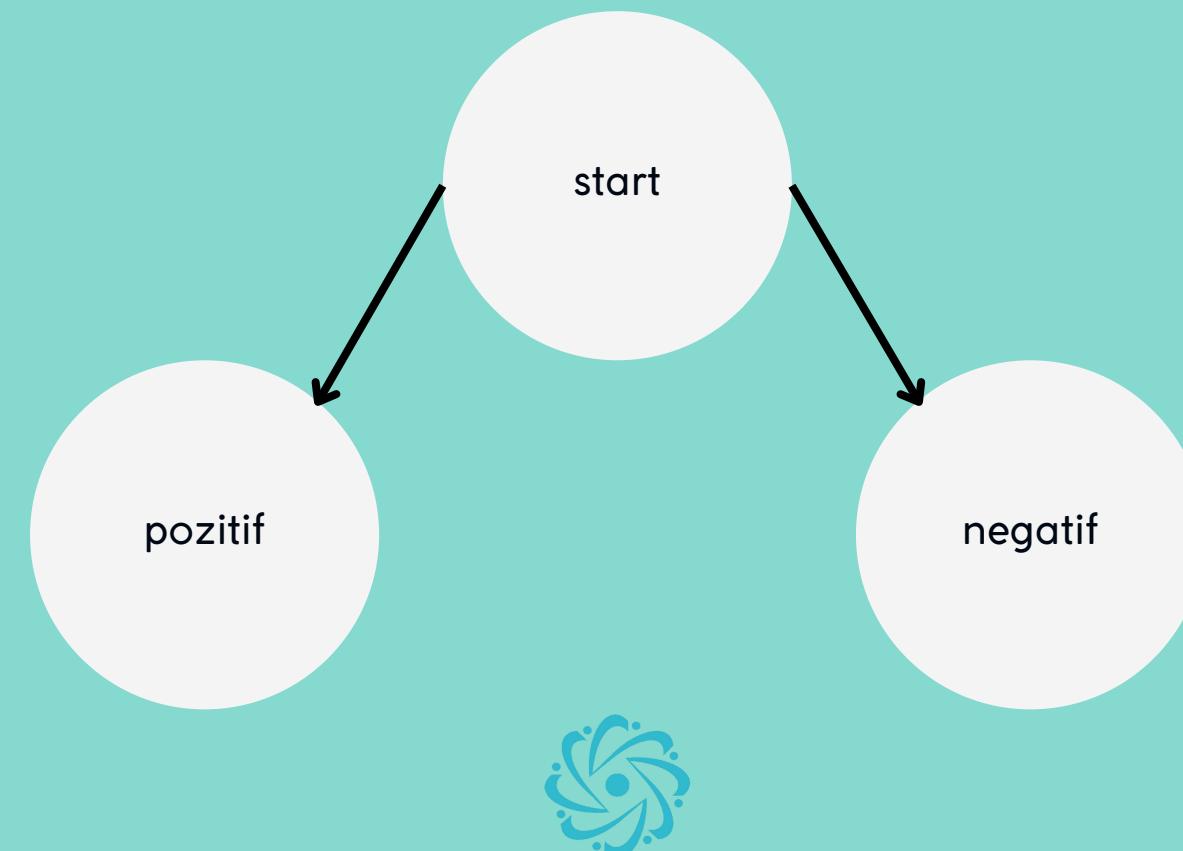
6. Virgülden Önce Pozitiflik Kontrolü (check_positive_beforeComma)

Virgülden önce pozitif bir bağlamın olup olmadığı kontrol edilir.

Kural:

positive_beforeComma: Pozitif bağlam varsa, FSM "Pozitif" durumuna geçer.

no_positive_beforeComma: Pozitif bağlam yoksa, FSM "check_end_with_degil" durumuna geçer.



FSMV3: Final Model

7. "Değil" ile Bitme Kontrolü (check_end_with_degil)

Cümplenin "değil" ile bitip bitmediği kontrol edilir.

Kural:

end_with_degil: Cümle "değil" ile bitiyorsa, FSM "Negatif" durumuna geçer.

no_end_with_degil: Cümle "değil" ile bitmiyorsa, FSM "check_conjunctions_word" durumuna geçer.

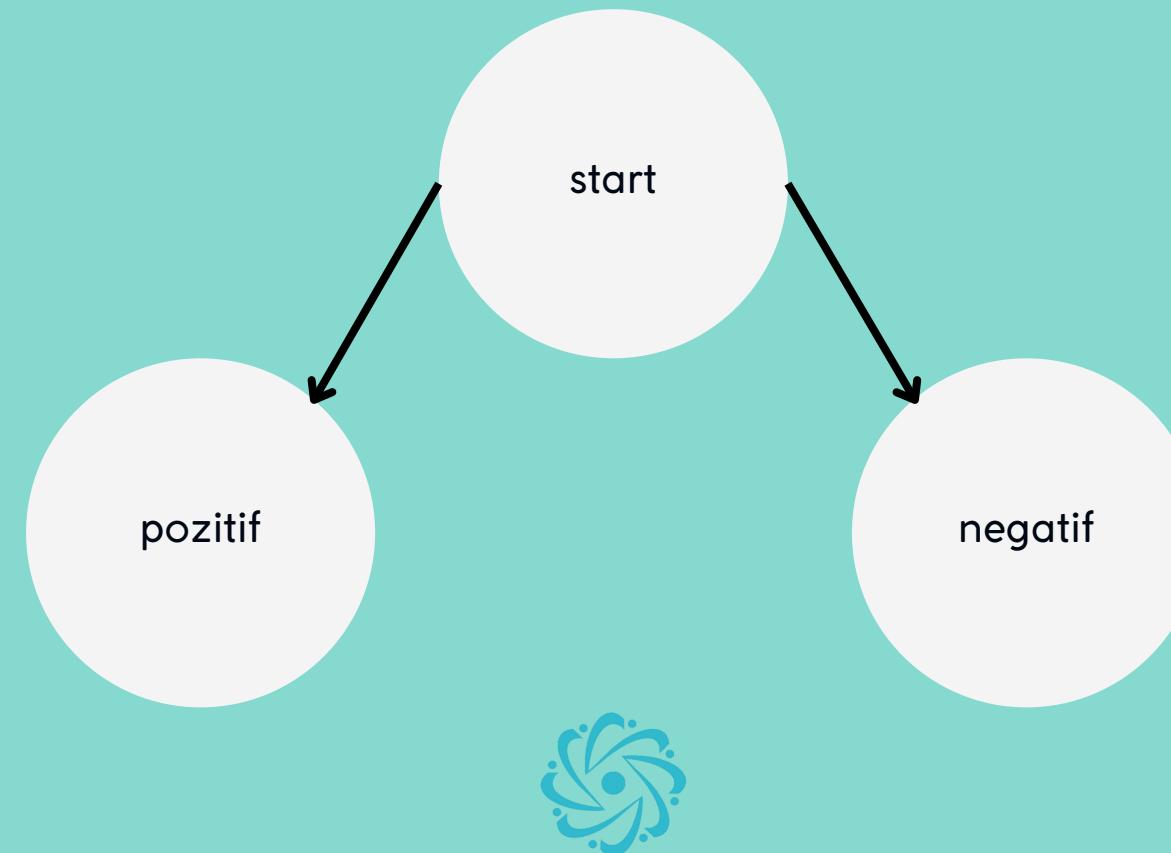
8. Bağlaç Kelime Kontrolü (check_conjunctions_word)

Cümlede bağlaç kelimelerle negatif bir bağlam oluşturulup oluşturulmadığını kontrol eder.

Kural:

conjunctions_word: Eğer bağlaçlar negatif bir anlam taşıyorsa, FSM "Negatif" durumuna geçer.

no_conjunctions_word: Böyle bir durum yoksa, FSM "check_hic_before_pos" durumuna geçer.



FSMV3: Final Model

9. "Hiç" ve Pozitif Bağlam Kontrolü (check_hic_before_pos)

"Hiç" kelimesinin pozitif bir bağlamda kullanılıp kullanılmadığını kontrol eder.

Kural:

hic_before_pos: Eğer "hic" pozitif bir bağlamı olumsuz yapıyorsa, FSM "Negatif" durumuna geçer.

no_hic_before_pos: Böyle bir durum yoksa, FSM "check_hic_before_neg" durumuna geçer.

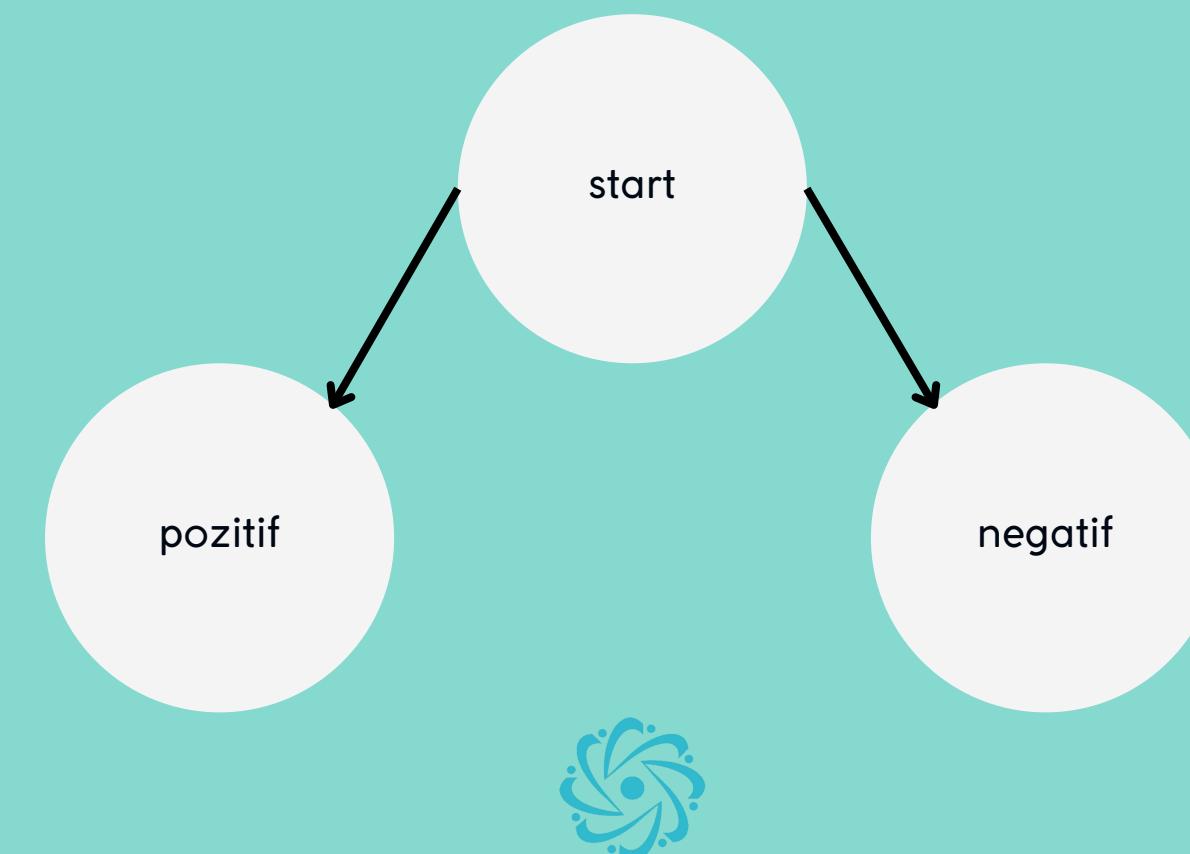
10. "Hiç" ve Negatif Bağlam Kontrolü (check_hic_before_neg)

"Hiç" kelimesinin negatif bir bağlamda kullanılmasını değerlendirdir.

Kural:

hic_before_neg: Eğer "hic" negatif bir bağlamı güçlendiriyorsa, FSM "Pozitif" durumuna geçer.

no_hic_before_neg: Böyle bir durum yoksa, FSM "resolve_tie" durumuna geçer.



11. Skor Tabanlı Karar (resolve_tie)

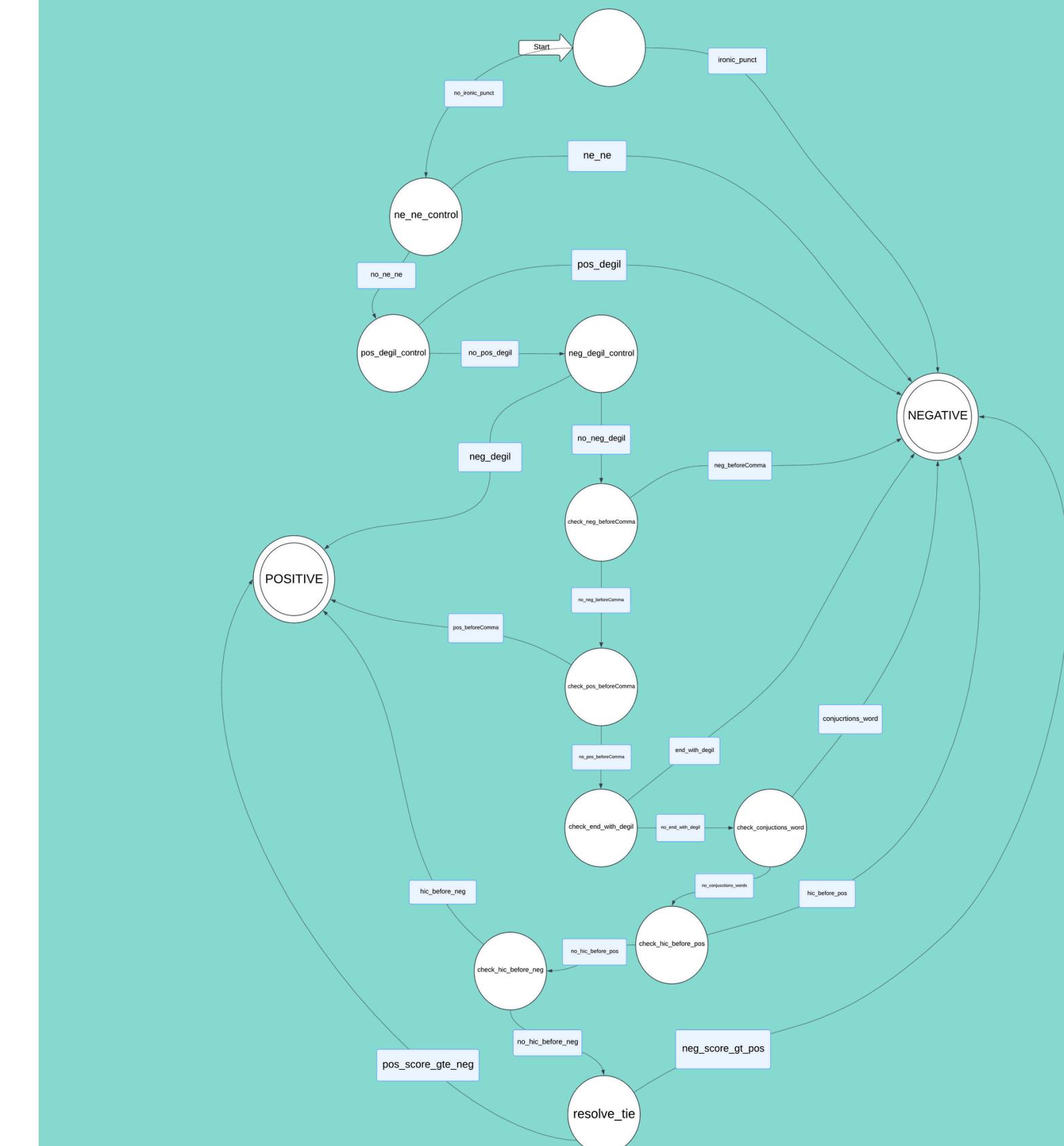
Pozitif ve negatif kelime skorları karşılaştırılarak nihai karar verilir.

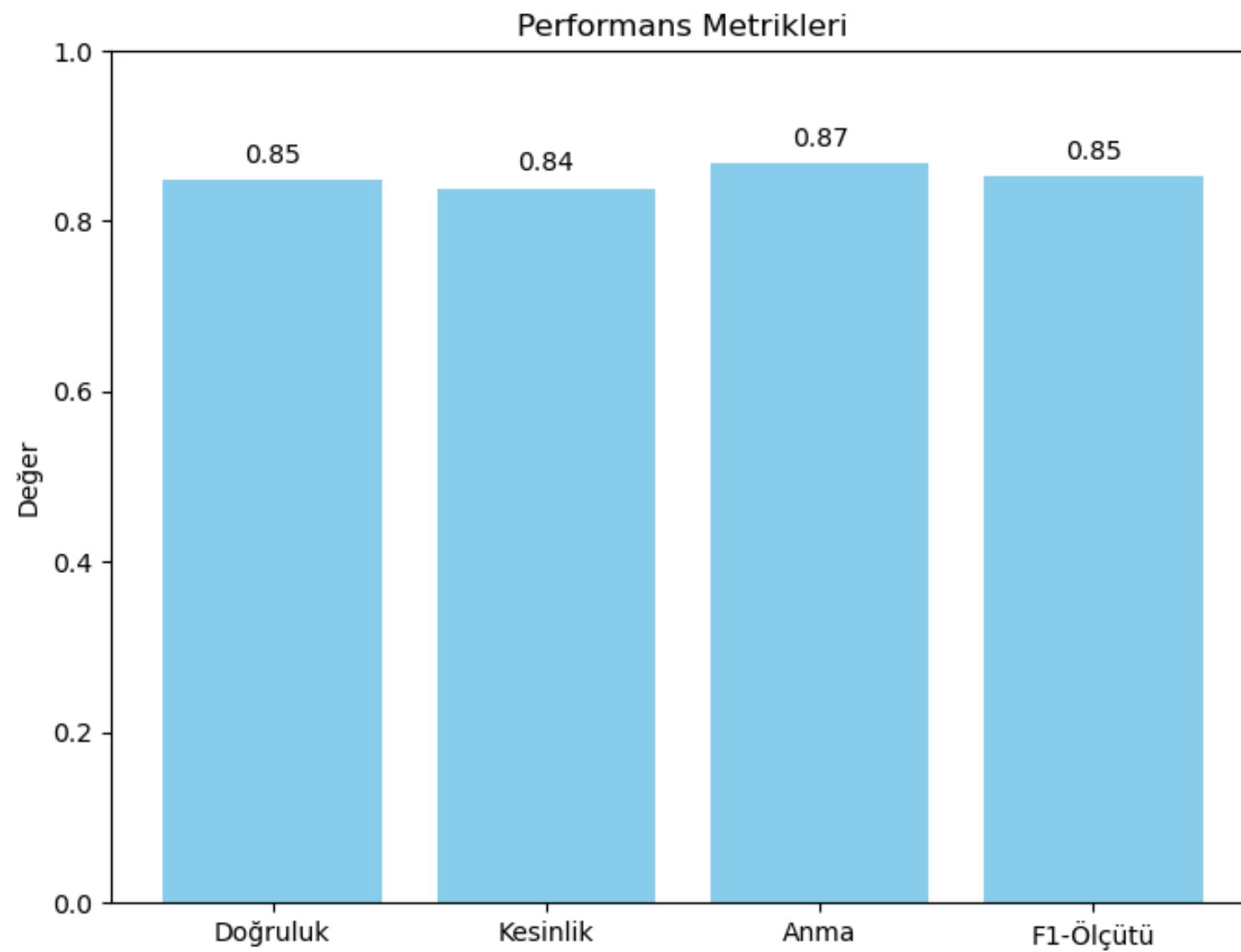
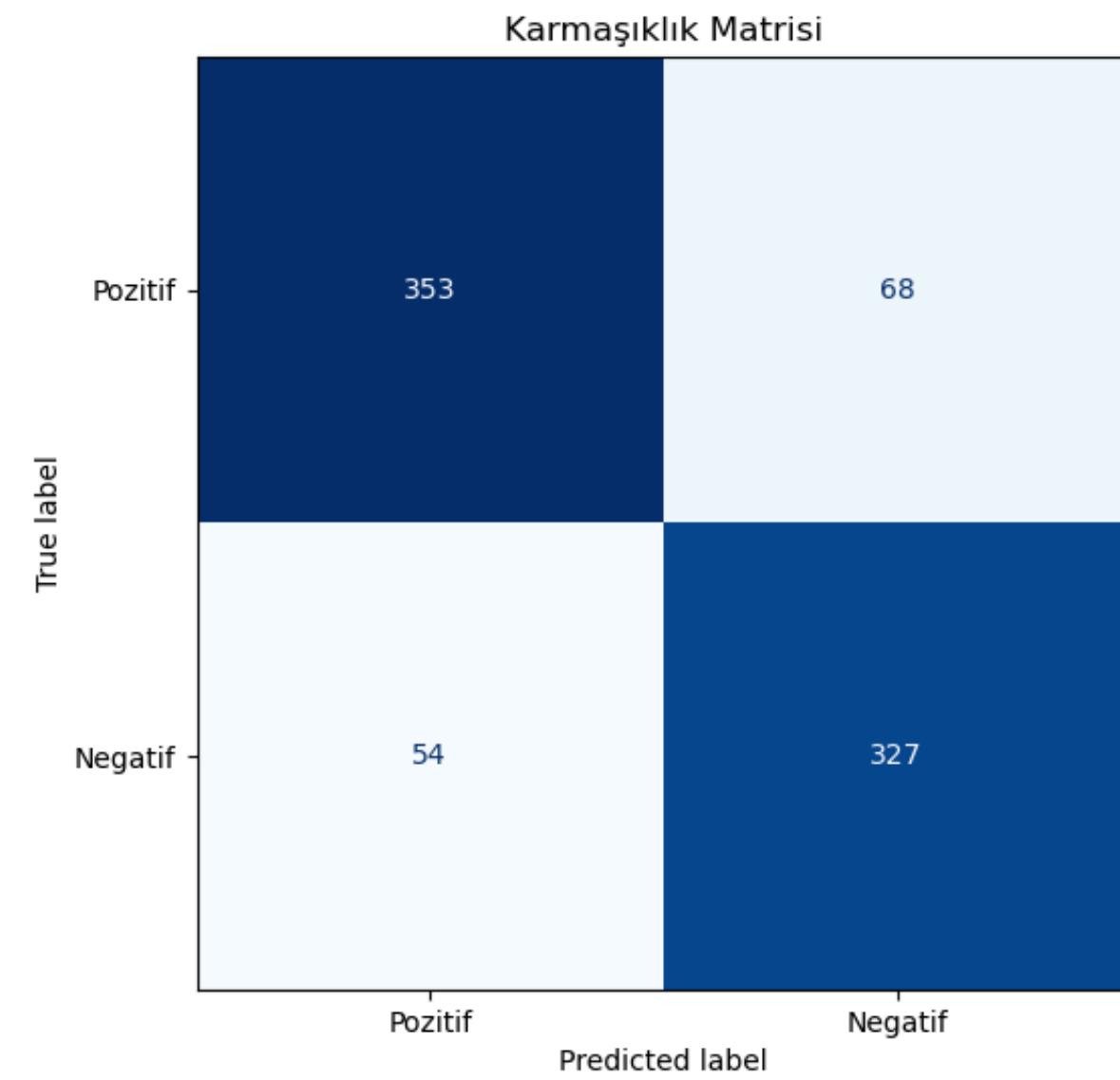
Kural:

positive_score_gte_negative: Pozitif skorlar negatif skorlara eşit ya da daha büyükse, FSM "Pozitif" durumuna geçer.

negative_score_gt_positive: Negatif skorlar daha büyükse, FSM "Negatif" durumuna geçer.

Bu yeni FSM modeli, önceki modellerin güçlü yönlerini korurken bağlaçlar ile bağlamı daha başarılı algılar

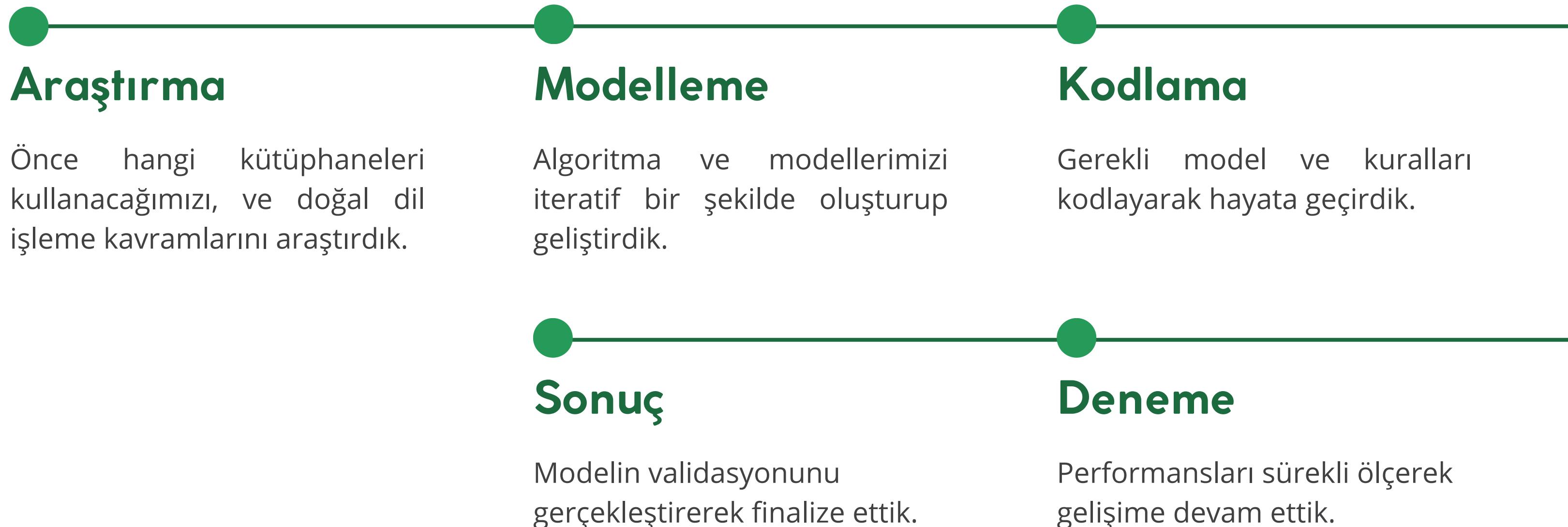




Doğruluk oranı %82'den %85'e yükselerek modelin genel doğruluğunun arttığını ortaya koymuştur. Kesinlik (%78 → %84) ve anma (%91 → %87) değerleri dengelenmiş, bu da modelin pozitif ve negatif sınıflamaları daha tutarlı bir şekilde yaptığına işaret etmektedir. F1-ölçütü ise %84'ten %85'e çıkararak kesinlik ve anma arasındaki uyumun güçlendiğini göstermektedir.



Uygulama



AKİŞ



Dosya Okuma ve Başlangıç Ayarları

Excel dosyası okunarak program başlatılır.

Zeyrek ve NLTK kütüphaneleri kullanılarak metin köklerine ve eklerine ayrılır.



Kuralların Uygulanması

Rules.py FSM'nin çalışma mantığını içeren fonksiyonları içerir.

Olumlu/Olumsuz kelimeler bağlaçlar ve bağlam gibi unsurlar değerlendirilir.



FSM'nin Çalıştırılması

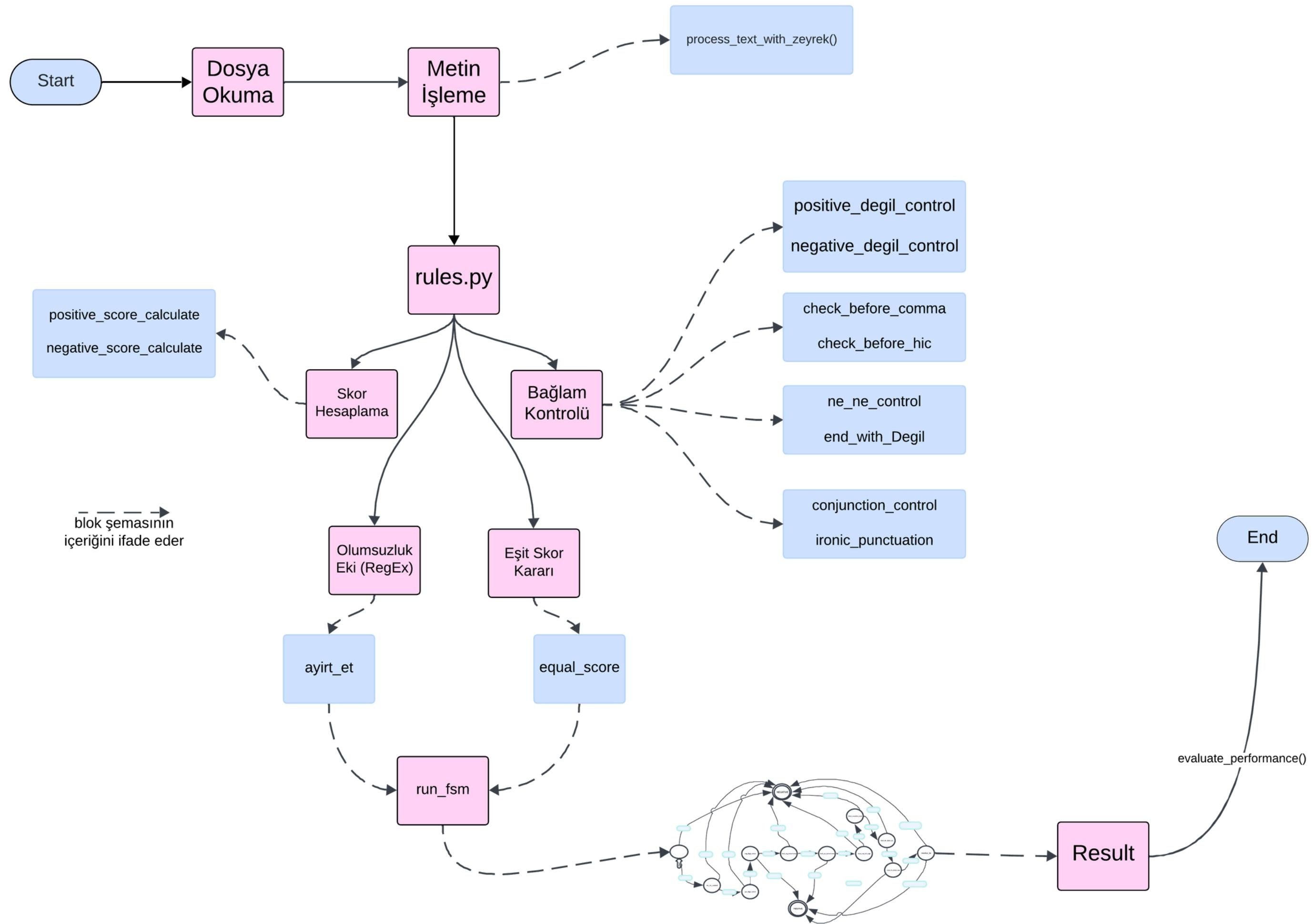
Kurallar soncunda FSM çalıştırılarak cumlenin duyu analizi gerçekleştirilir.



Performans Metrikleri

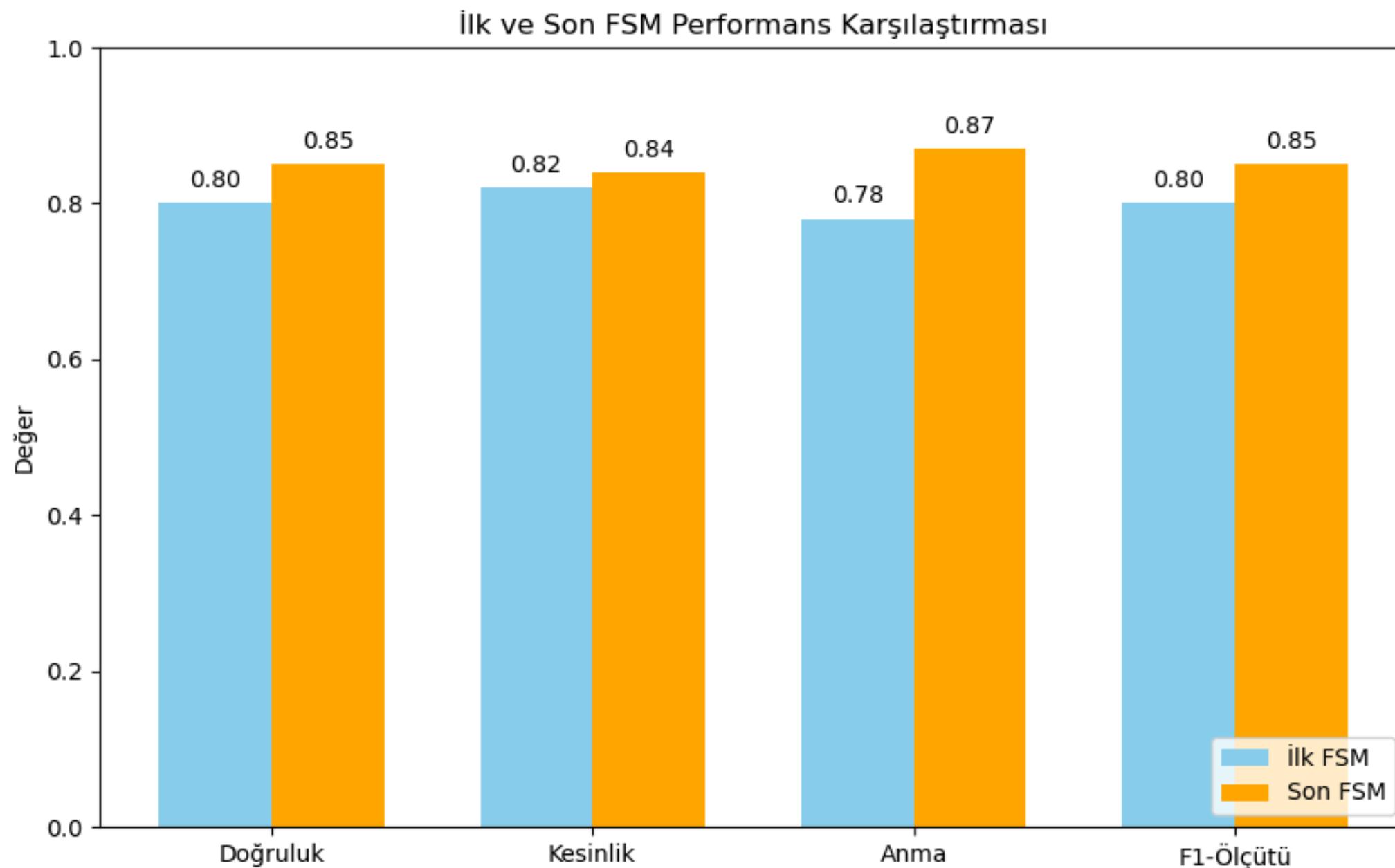
FSM'in tahminleri, gerçek sınıflarla karşılaştırılarak doğruluk, kesinlik, anma ve F1-ölcütü gibi metrikler hesaplanır.





Testler

FSMv2 ve FSMv3'ün Karşılaştırılması



Peki Yanlılık?

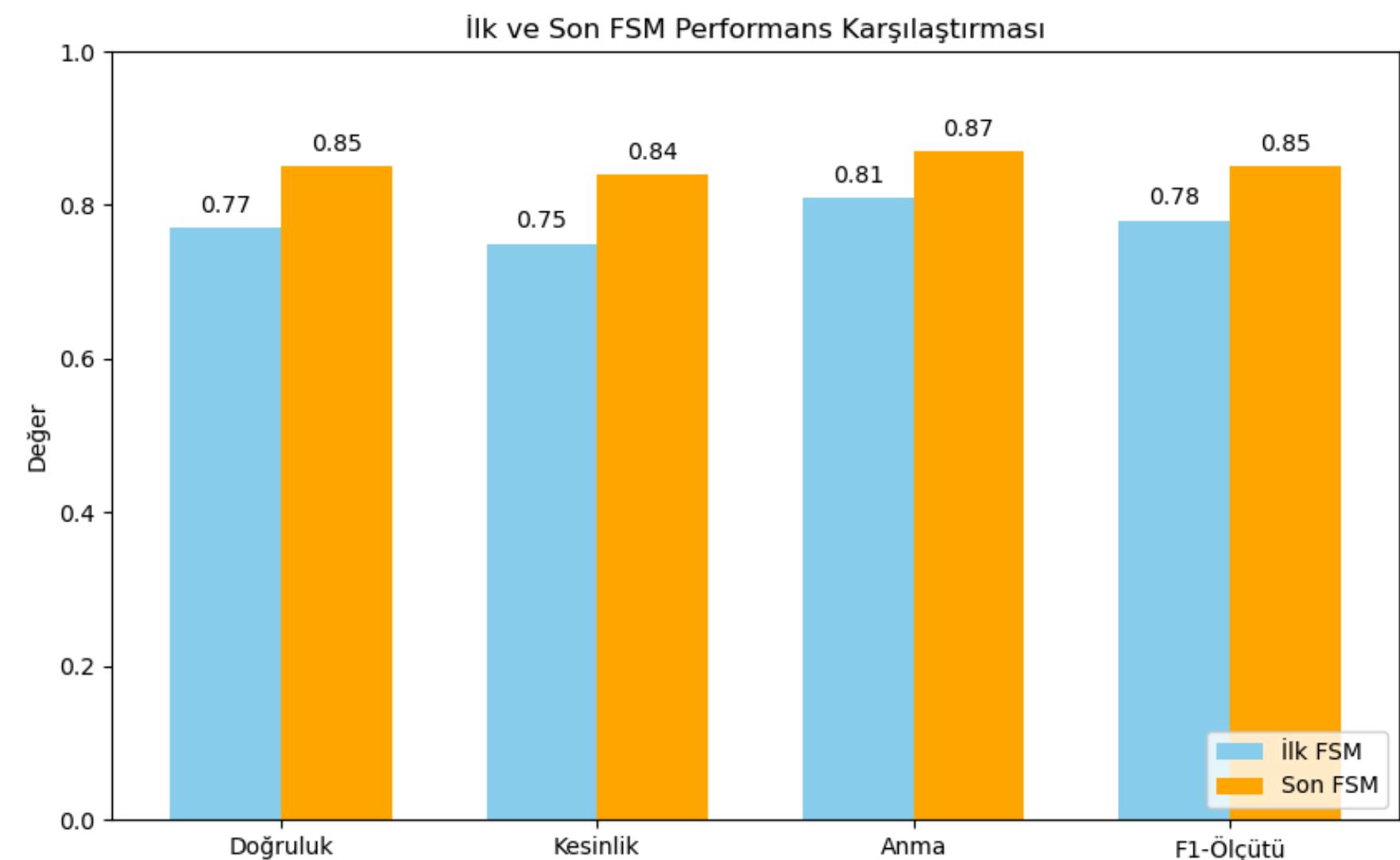


Model Validasyonu

Dengeli 200 gözlemlilik veri seti

FSMV3
Doğru Pozitif (DP): 88 **Doğruluk:** %87
Yanlış Pozitif (YP): 17 **Kesinlik:** %84
Yanlış Negatif (YN): 9 **Anma:** %91
Doğru Negatif (DN): 81 **F-1 Ölçütü:** %87

FSMV2
Doğru Pozitif (DP): 79 **Doğruluk:** %77
Yanlış Pozitif (YP): 26 **Kesinlik:** %75
Yanlış Negatif (YN): 18 **Anma:** %81
Doğru Negatif (DN): 72 **F-1 Ölçütü:** %78

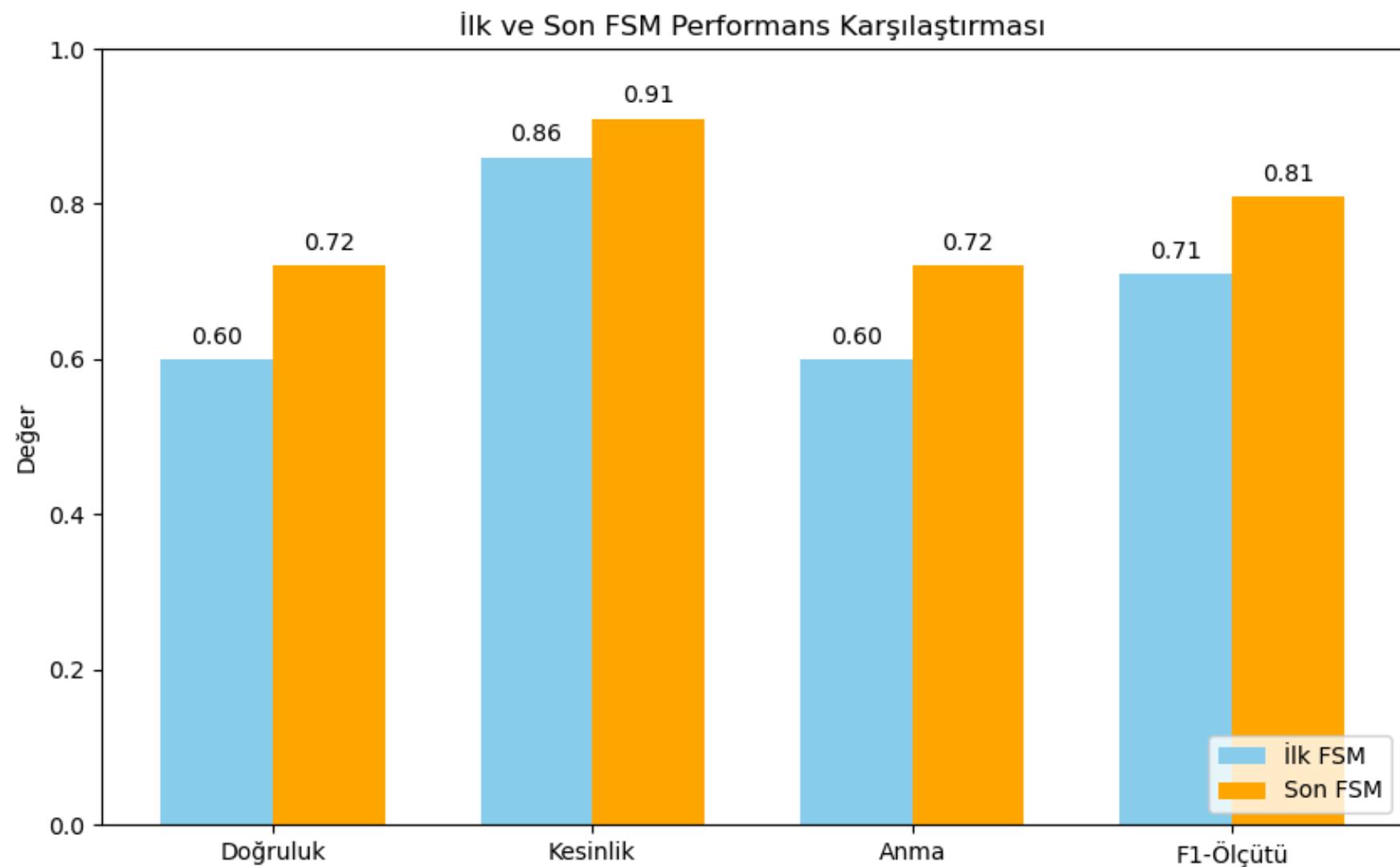


Model Validasyonu

Pozitif Ağırlıklı 100 gözlemlilik
veri seti (%80 pozitif)

FSMV3
Doğru Pozitif (DP): 58 Doğruluk: %72
Yanlış Pozitif (YP): 6 Kesinlik: %91
Yanlış Negatif (YN): 22 Anma: %72
Doğru Negatif (DN): 14 F-1 Ölçütü: %81

FSMV2
Doğru Pozitif (DP): 48 Doğruluk: %60
Yanlış Pozitif (YP): 8 Kesinlik: %86
Yanlış Negatif (YN): 32 Anma: %60
Doğru Negatif (DN): 12 F-1 Ölçütü: %71

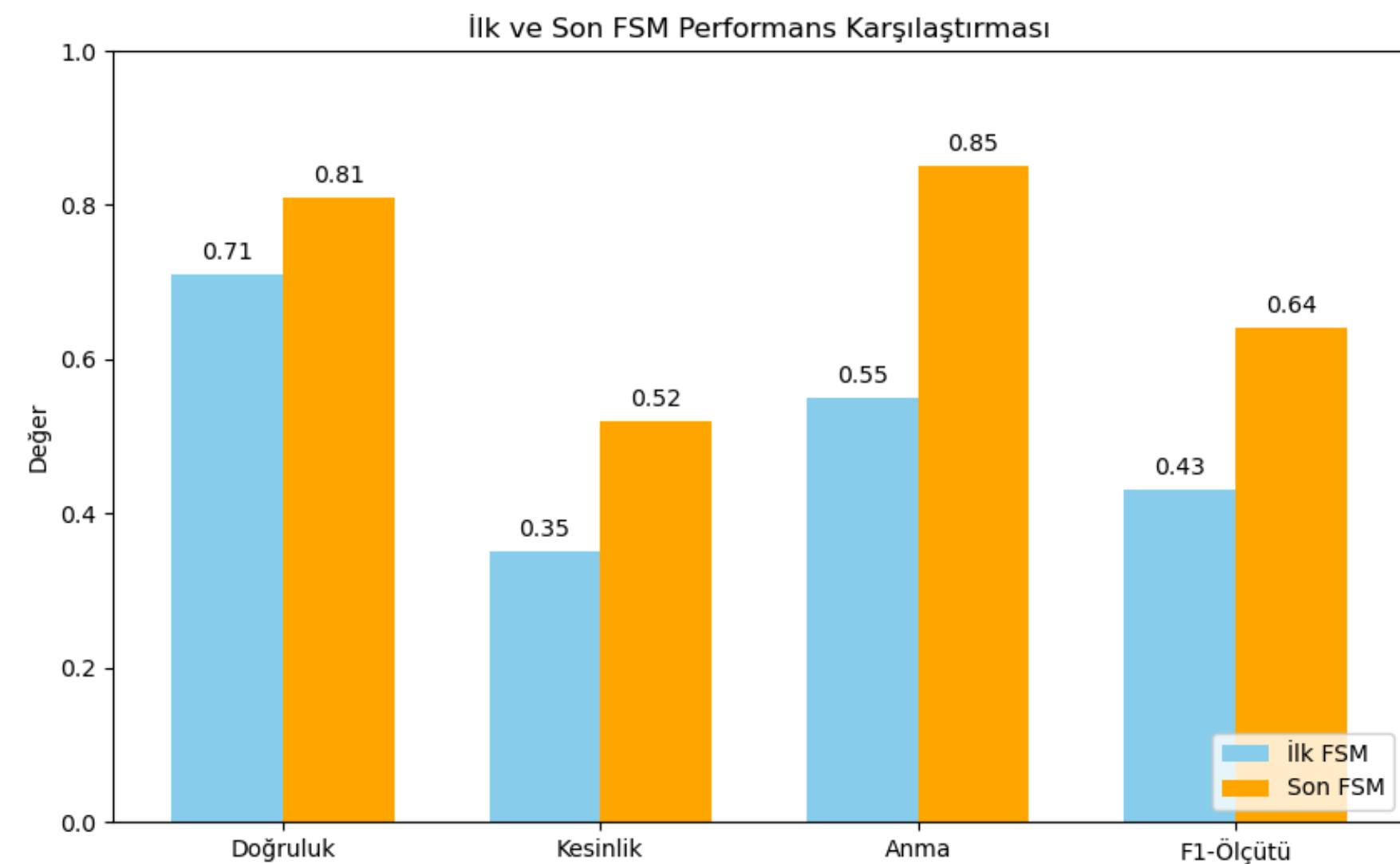


Model Validasyonu

Negatif Ağırlıklı 100 Gözlemlilik
Veri Seti (%80 Negatif)

FSMV3
Doğru Pozitif (DP): 17 Doğruluk: %80
Yanlış Pozitif (YP): 17 Kesinlik: %50
Yanlış Negatif (YN): 3 Anma: %85
Doğru Negatif (DN): 63 F-1 Ölçütü: %63

FSMV2
Doğru Pozitif (DP): 11 Doğruluk: %71
Yanlış Pozitif (YP): 20 Kesinlik: %35
Yanlış Negatif (YN): 9 Anma: %55
Doğru Negatif (DN): 60 F-1 Ölçütü: %43



Tesekkürler.

