



Mühendislik ve Doğa Bilimleri Fakültesi
Bilgisayar Mühendisliği Bölümü

BİL485 Derin Öğrenme

LSTM Tabanlı Günlük Sıcaklık Tahmini:
Melbourne Verisi Üzerine Bir Zaman Serisi Analizi

Hazırlayanlar:

22120205043 Şevval ŞAHİN

Öğretim Görevlisi: Dr. Öğr. Üyesi İshak Dölek

İçindekiler Tablosu

1. PROJE KONUSU	4
1.1 Seçilme Gerekçesi	4
1.2 Alanın Önemi	5
2. VERİ SETİNİN BELİRLENMESİ	5
3. UYGULANACAK YÖNTEM ve LİTERATÜR KARŞILAŞTIRMASI	6
3.1. Uygulanacak Yöntem	6
3.2. Literatür Karşılaştırması.....	6
4. MODEL EĞİTİMİ	7
5. MODEL DEĞERLENDİRMESİ	8
6. SONUÇ	9
7. KAYNAKÇA	11
8. EKLER	12
Ek A: Proje Kaynak Kodları	12
Ek B: Kullanılan Veri Seti.....	12
Ek C: Model Servisleme ve Kullanıcı Arayüzü	12

ŞEKİL LİSTESİ

Şekil 1 Eğitim Kayıp Grafiği	8
Şekil 2 Gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması	9
Şekil 3 Gradio kullanıcı arayüzü.....	12

1. PROJE KONUSU

Bu proje kapsamında, günlük minimum sıcaklık değerlerinin zaman serisi analizi yoluyla tahmin edilmesi amaçlanmıştır. Çalışmada, derin öğrenme alanında yaygın olarak kullanılan Uzun Kısa Süreli Bellek (Long Short-Term Memory – LSTM) tabanlı bir sinir ağı modeli geliştirilerek, geçmiş günlere ait sıcaklık verilerinden yararlanılarak bir sonraki günün minimum sıcaklık değeri tahmin edilmektedir.

Meteorolojik veriler doğası gereği zamana bağlıdır ve geçmiş gözlemler ile gelecekteki değerler arasında güçlü ilişkiler barındırır. Bu nedenle söz konusu problem, zaman serisi tahmini kapsamında ele alınmıştır. Projede kullanılan LSTM modeli, geleneksel yöntemlerin yetersiz kaldığı doğrusal olmayan ilişkileri öğrenebilme yeteneğine sahip olması nedeniyle tercih edilmiştir. Geliştirilen sistem, hem akademik hem de pratik açıdan zaman serisi tahmin problemlerine uygulanabilir bir çözüm sunmaktadır.

1.1 Seçilme Gerekçesi

Zaman serisi tahmini; meteoroloji, enerji yönetimi, finans ve sağlık gibi birçok alanda kritik öneme sahip bir problemdir. Özellikle sıcaklık tahminleri, tarımsal faaliyetlerin planlanması, enerji tüketiminin öngörülmesi ve iklim değişikliği analizleri açısından büyük önem taşımaktadır. Günlük minimum sıcaklık değerlerinin doğru şekilde tahmin edilmesi, don olaylarının öngörülmesi ve iklimsel risklerin azaltılması gibi uygulamalarda doğrudan kullanılabilmektedir.

Geleneksel istatistiksel yöntemler, zaman serisi verilerini analiz etmede uzun yıllardır kullanılmasına rağmen, genellikle doğrusal varsayımlara dayanmakta ve karmaşık örüntüleri yakalamakta sınırlı kalmaktadır. Gerçek dünya verileri ise çoğu zaman gürültülü, doğrusal olmayan ve uzun dönemli bağımlılıklar içeren yapılar sergilemektedir. Bu durum, daha esnek ve güçlü modelleme yaklaşımlarına olan ihtiyacı artırmaktadır.

Bu projede derin öğrenme tabanlı bir yaklaşım benimsenmesinin temel gerekçesi, LSTM modellerinin uzun dönemli zaman bağımlılıklarını öğrenebilme yeteneğidir. LSTM mimarisi, klasik yapay sinir ağlarının aksine, geçmiş bilgiyi hücresel bellek yapısı aracılığıyla koruyabilmekte ve bu sayede zaman serilerindeki uzun vadeli ilişkileri etkili bir şekilde modelleyebilmektedir. Bu özellikleri nedeniyle LSTM, sıcaklık tahmini gibi zaman serisi problemleri için uygun ve güçlü bir yöntem olarak değerlendirilmiştir.

Ayrıca, proje kapsamında geliştirilen modelin yalnızca teorik olarak değil, aynı zamanda kullanıcı etkileşimli bir arayüz (Gradio) aracılığıyla pratik olarak da test edilebilir olması hedeflenmiştir. Bu sayede modelin gerçek dünyada nasıl kullanılabileceği gösterilmiş ve uygulama odaklı bir yaklaşım benimsenmiştir.

1.2 Alanın Önemi

Zaman serisi tahmini, geçmişte gözlemlenen verilerden yararlanarak geleceğe yönelik öngörülerde bulunmayı amaçlayan ve birçok uygulama alanında önemli bir yere sahip olan bir yöntemdir. Özellikle meteorolojik verilerin analizi, tarım, enerji yönetimi ve iklim çalışmaları açısından kritik öneme sahiptir. Günlük minimum sıcaklık değerlerinin doğru şekilde tahmin edilmesi, çevresel risklerin azaltılması ve kaynak planlamasının daha verimli yapılabilmesi açısından büyük katkı sağlamaktadır.

Sıcaklık verileri genellikle doğrusal olmayan ve zamana bağlı bir yapı sergilediğinden, klasik istatistiksel yöntemler bu tür veriler üzerinde sınırlı başarı göstermektedir. Bu nedenle, zaman serilerindeki karmaşık ilişkileri öğrenebilen derin öğrenme tabanlı yaklaşımlar ön plana çıkmaktadır. Özellikle LSTM modelleri, uzun dönemli bağımlılıkları öğrenebilme yetenekleri sayesinde sıcaklık tahmini gibi problemlerde etkili sonuçlar sunmaktadır.

Bu bağlamda, zaman serisi analizi ve derin öğrenme yöntemleri, hem akademik araştırmalar hem de gerçek dünya uygulamaları açısından giderek artan bir öneme sahiptir. Bu projede ele alınan problem, söz konusu alanın önemini ortaya koyan somut bir uygulama örneği sunmaktadır.

2. VERİ SETİNİN BELİRLENMESİ

Bu çalışmada kullanılan veri seti, Kaggle platformunda paylaşılan “Daily Minimum Temperatures in Melbourne” veri setidir. Veri seti, Avustralya’nın Melbourne şehrinde 1981–1990 yılları arasında ölçülen günlük minimum sıcaklık değerlerini içermektedir. Toplamda yaklaşık on yıllık bir zaman aralığını kapsayan bu veri seti, zaman serisi tahmin problemleri için yeterli uzunlukta ve düzenli bir yapıya sahiptir.

Veri seti iki temel sütundan oluşmaktadır: tarih bilgisi ve ilgili güne ait minimum sıcaklık değeri. Tarih bilgisi zaman serisinin kronolojik yapısını temsil ederken, sıcaklık değeri model tarafından tahmin edilmesi hedeflenen değişkendir. Çalışma kapsamında yalnızca sıcaklık değerleri kullanılmış, tarih bilgisi ise verinin sıralı yapısının korunması amacıyla değerlendirilmiştir.

Model eğitime geçmeden önce veri seti üzerinde ön işleme adımları uygulanmıştır. Eksik veya geçersiz değerler veri setinden çıkarılmış, sıcaklık değerleri sayısal formata dönüştürülmüştür. Ayrıca, modelin daha kararlı ve hızlı öğrenebilmesi için sıcaklık verileri Min-Max ölçekleme yöntemi kullanılarak 0–1 aralığına normalize edilmiştir.

Zaman serisi verisinin LSTM modeline uygun hale getirilebilmesi amacıyla sliding window yaklaşımı kullanılmıştır. Bu yaklaşımda, önceki 7 güne ait sıcaklık değerleri giriş verisi olarak kullanılmış ve bir sonraki günün minimum sıcaklık değeri hedef değişken olarak belirlenmiştir. Veri seti, zaman bağımlılığı korunacak şekilde eğitim ve test kümelerine ayrılmıştır.

Seilen veri seti, hem gerek dnya verisi olması hem de zaman serisi yapısına uygunluęu nedeniyle bu proje kapsamında uygulanacak derin ęrenme yaklaşımlarını deęerlendirmek iin uygun bir rnek sunmaktadır.

3. UYGULANACAK YNTEM ve LİTERATR KARŞILAŞTIRMASI

3.1. Uygulanacak Yntem

Bu projede, gnlk minimum sıcaklık deęerlerinin tahmin edilmesi amacıyla Uzun Kısa Sreli Bellek (Long Short-Term Memory – LSTM) tabanlı bir derin ęrenme modeli kullanılmıştır. LSTM, tekrarlayan sinir aęları (Recurrent Neural Networks – RNN) ailesine ait olup, zaman serisi verilerinde grlen uzun dnemli baęımlılıkları ęrenebilme yeteneęi ile ne ıkmaktadır.

Geleneksel RNN yapıları, uzun diziler zerinde alışırken gradyan kaybolması ve patlaması gibi problemlerle karşılaşılabilmektedir. LSTM mimarisi ise hcresel bellek yapısı ve kapı (gate) mekanizmaları sayesinde bu problemleri byk lde ortadan kaldırmakta ve gemiş bilginin daha uzun sre korunmasını saęlamaktadır. Bu zellik, sıcaklık gibi gemiş deęerlerin geleceęi doęrudan etkiledięi zaman serisi problemleri iin LSTM modellerini uygun hale getirmektedir.

Bu alışma kapsamında geliştirilen LSTM modeli, nceki 7 gne ait sıcaklık deęerlerini girdi olarak almakta ve bir sonraki gnn minimum sıcaklık deęerini tahmin etmektedir. Model mimarisi, tek katmanlı bir LSTM ve bunu takip eden tam baęlantılı (Dense) bir katmandan oluşmaktadır. Model, Adam optimizasyon algoritması ve ortalama kare hata (Mean Squared Error – MSE) kayıp fonksiyonu kullanılarak eęitilmiştir. Bu yapı, modelin hem yeterli ęrenme kapasitesine sahip olmasını hem de aşırı ęrenmenin nne geilmesini saęlamaktadır.

3.2. Literatr Karşılaştırması

Zaman serisi tahmini alanında literatrde farklı yntemler yaygın olarak kullanılmaktadır. Klasik istatistiksel yaklaşımlar arasında ARIMA, SARIMA ve Holt–Winters gibi modeller yer almaktadır. Bu yntemler, zellikle duraęan ve kısa dnemli zaman serilerinde başarılı sonuçlar verebilmekle birlikte, doęrusal varsayımlara dayanmaları nedeniyle karmaşık ve doęrusal olmayan veri yapılarında sınırlı performans gstermektedir.

Makine ęrenmesi alanındaki gelişmelerle birlikte, Destek Vektr Makineleri (SVM), Yapay Sinir Aęları (ANN) ve Rastgele Ormanlar (Random Forest) gibi yntemler zaman serisi tahmininde kullanılmaya başlanmıştır. Bu yntemler, klasik istatistiksel modellere kıyasla daha esnek olmakla birlikte, zaman baęımlılıklarını doęrudan modelleme konusunda bazı kısıtlamalara sahiptir.

Son yıllarda yapılan çalışmalarda, LSTM ve GRU gibi derin öğrenme tabanlı tekrarlayan sinir ağlarının, sıcaklık tahmini ve benzeri zaman serisi problemlerinde daha yüksek doğruluk sağladığı gösterilmiştir. LSTM modelleri, uzun dönemli bağımlılıkları öğrenebilme yetenekleri sayesinde, özellikle meteorolojik veriler üzerinde klasik ve geleneksel makine öğrenmesi yöntemlerine kıyasla üstün performans sergilemektedir.

Bu proje kapsamında LSTM tabanlı bir yaklaşım tercih edilerek, literatürde rapor edilen başarılı uygulamalarla uyumlu bir yöntem izlenmiştir. Elde edilen sonuçlar, LSTM mimarisinin günlük minimum sıcaklık tahmini probleminde etkili bir yöntem olduğunu ve literatürdeki benzer çalışmalarla tutarlı sonuçlar ürettiğini göstermektedir.

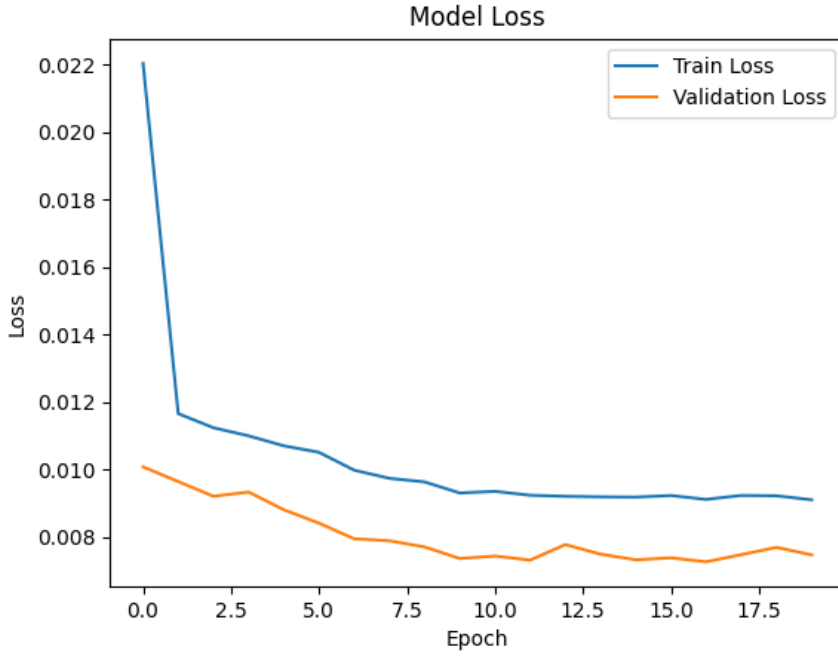
4. MODEL EĞİTİMİ

Bu çalışmada geliştirilen LSTM tabanlı model, günlük minimum sıcaklık değerlerinin tahmin edilmesi amacıyla eğitilmiştir. Model eğitime başlamadan önce veri seti, zaman serisi yapısı korunacak şekilde eğitim ve test kümelerine ayrılmıştır. Veri setinin %80'i modelin eğitimi için, %20'si ise modelin performansının değerlendirilmesi için kullanılmıştır. Bu ayırım sırasında kronolojik sıralama korunarak geleceğe ait verilerin eğitim sürecine dâhil edilmemesine dikkat edilmiştir.

Modelin eğitimi sırasında, her bir giriş örneği önceki 7 güne ait sıcaklık değerlerinden oluşmaktadır. Bu değerler, LSTM modeline uygun olacak şekilde üç boyutlu bir yapıya dönüştürülmüştür. Eğitim sürecinde model, Adam optimizasyon algoritması kullanılarak 20 epoch boyunca eğitilmiştir. Kayıp fonksiyonu olarak ortalama kare hata (Mean Squared Error – MSE) tercih edilmiştir. Bu kayıp fonksiyonu, tahmin edilen değerler ile gerçek değerler arasındaki farkı ölçerek modelin öğrenme sürecini yönlendirmektedir.

Eğitim sürecinde, modelin yalnızca eğitim verisi üzerinde değil, doğrulama verisi üzerinde de performansı izlenmiştir. Bu sayede modelin aşırı öğrenme (overfitting) eğilimi gösterip göstermediği kontrol edilmiştir. Eğitim ve doğrulama kayıplarının epochlara göre değişimi grafiksel olarak incelenmiş ve modelin kararlı bir öğrenme süreci gerçekleştirdiği gözlemlenmiştir.

Eğitim tamamlandıktan sonra model, ileride tekrar kullanılabilmesi ve servis edilebilmesi amacıyla dosya olarak kaydedilmiştir. Bu aşama, modelin yalnızca teorik olarak değil, aynı zamanda pratik uygulamalarda da kullanılabilir olduğunu göstermektedir.



Şekil 1 Eğitim Kayıp Grafiği

Şekil 1’de modelin eğitim sürecine ait eğitim ve doğrulama kayıp değerlerinin epochlara göre değişimi gösterilmektedir. Eğitim sürecinin ilk epochlarında kayıp değerlerinde hızlı bir düşüş gözlemlenirken, ilerleyen epochlarda bu düşüşün daha dengeli bir şekilde devam ettiği görülmektedir. Eğitim ve doğrulama kayıplarının birbirine yakın seyretmesi, modelin kararlı bir öğrenme süreci gerçekleştirdiğini ve aşırı öğrenme eğilimi göstermediğini ortaya koymaktadır.

5. MODEL DEĞERLENDİRMESİ

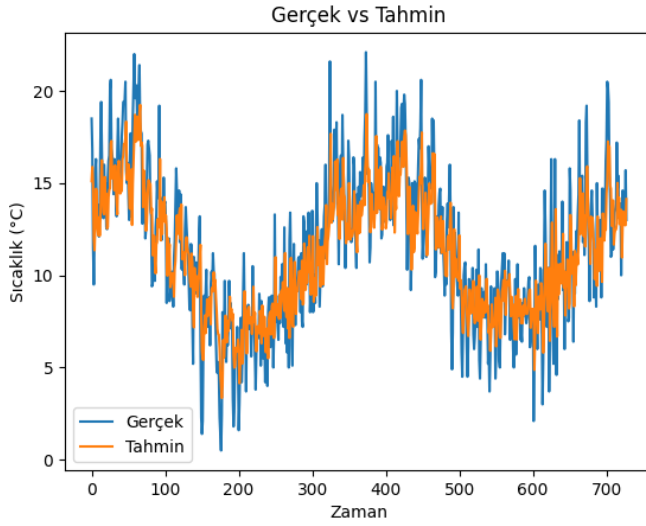
Modelin performansını değerlendirmek amacıyla, eğitim sürecinde kullanılmayan test veri kümesi üzerinde tahminler yapılmıştır. Test verisi, modelin daha önce görmediği verilerden oluştuğu için elde edilen sonuçlar modelin genelleme yeteneğini yansıtmaktadır. Bu aşamada, LSTM modelinin ürettiği tahmin değerleri gerçek sıcaklık değerleri ile karşılaştırılmıştır.

Değerlendirme sürecinde hata metriği olarak kök ortalama kare hata (Root Mean Squared Error – RMSE) kullanılmıştır. RMSE metriği, tahmin edilen değerler ile gerçek değerler arasındaki farkların karesinin ortalamasının karekökünü alarak modelin hata düzeyini ölçmektedir. Bu metrik, sıcaklık gibi sürekli değerler içeren regresyon problemlerinde yaygın olarak tercih edilmektedir. Elde edilen RMSE değeri, modelin günlük minimum sıcaklık değerlerini makul bir hata payı ile tahmin edebildiğini göstermektedir.

Modelin tahmin performansı ayrıca grafiksel olarak da incelenmiştir. Gerçek sıcaklık değerleri ile model tarafından üretilen tahmin değerleri aynı grafik üzerinde gösterilerek karşılaştırılmıştır. Bu karşılaştırma, modelin genel eğilimi yakalayabildiğini ve sıcaklık

değişimlerini büyük ölçüde takip edebildiğini ortaya koymaktadır. Bazı dönemlerde küçük sapmalar gözlemlense de, modelin genel olarak gerçek değerlere yakın tahminler ürettiği görülmüştür.

Elde edilen sonuçlar, LSTM tabanlı modelin zaman serisi yapısına sahip sıcaklık verileri üzerinde etkili bir şekilde çalıştığını ve günlük minimum sıcaklık tahmini problemi için uygun bir yaklaşım sunduğunu göstermektedir.



Şekil 2 Gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

Şekil 2’de test veri kümesi üzerinde gerçek sıcaklık değerleri ile LSTM modeli tarafından üretilen tahmin değerleri karşılaştırmalı olarak gösterilmektedir. Grafik incelendiğinde, modelin sıcaklık verisinin genel eğilimini başarılı bir şekilde takip edebildiği görülmektedir. Bazı dönemlerde kısa süreli sapmalar gözlemlense de, tahmin değerlerinin büyük ölçüde gerçek değerlere yakın olduğu ve modelin zaman serisi yapısını öğrenebildiği anlaşılmaktadır.

6. SONUÇ

Bu proje kapsamında, günlük minimum sıcaklık değerlerinin tahmin edilmesi amacıyla LSTM tabanlı bir derin öğrenme modeli geliştirilmiştir. Gerçek dünya verilerinden oluşan bir zaman serisi veri seti kullanılarak model eğitilmiş ve değerlendirilmiştir. Uygulanan ön işleme adımları ve seçilen model mimarisi sayesinde modelin zaman serisi yapısını öğrenebildiği gözlemlenmiştir.

Elde edilen sonuçlar, LSTM modelinin geçmiş günlere ait sıcaklık değerlerini kullanarak bir sonraki günün minimum sıcaklık değerini makul bir doğrulukla tahmin edebildiğini göstermektedir. Eğitim ve doğrulama süreçlerinde elde edilen kayıp değerleri, modelin kararlı bir öğrenme süreci gerçekleştirdiğini ve aşırı öğrenme eğilimi göstermediğini ortaya koymaktadır. Test verisi üzerinde hesaplanan hata metriği ise modelin genelleme yeteneğinin yeterli düzeyde olduğunu desteklemektedir.

Ayrıca, geliştirilen modelin bir kullanıcı arayüzü aracılığıyla servis edilmesi, çalışmanın yalnızca teorik değil, aynı zamanda uygulamaya yönelik bir nitelik taşıdığını göstermektedir. Bu yönüyle proje, zaman serisi tahmini problemlerinde derin öğrenme yöntemlerinin etkinliğini ortaya koyan başarılı bir örnek sunmaktadır.

Gelecek çalışmalarda, farklı pencere uzunlukları, daha derin LSTM mimarileri veya alternatif derin öğrenme modelleri kullanılarak tahmin performansının artırılması hedeflenebilir. Bunun yanı sıra, farklı meteorolojik değişkenlerin modele dâhil edilmesiyle daha kapsamlı tahmin sistemleri geliştirilebilir.

7. KAYNAKÇA

1. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
3. Brownlee, J. (2017). *Deep learning for time series forecasting*. Machine Learning Mastery.
4. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. Wiley.
5. Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Manning Publications.
6. Kaggle. (2018). Daily minimum temperatures in Melbourne. <https://www.kaggle.com/datasets/paulbrabban/daily-minimum-temperatures-in-melbourne>

8. EKLER

Ek A: Proje Kaynak Kodları

Bu çalışmada geliştirilen derin öğrenme modeli ve projeye ait tüm kaynak kodları, sürüm kontrolü ve düzenli dokümantasyon amacıyla GitHub platformu üzerinde paylaşılmıştır. GitHub deposu içerisinde model mimarisi, eğitim süreci, veri ön işleme adımları ve modelin servis edilmesine yönelik kodlar yer almaktadır. Ayrıca proje dosya yapısı, çalışmanın tekrar üretilebilirliğini sağlayacak şekilde düzenlenmiştir.

GitHub bağlantısı:

<https://github.com/sevvalsaahinn/lstm-daily-temperature-prediction/>

Ek B: Kullanılan Veri Seti

Bu projede kullanılan veri seti, Kaggle platformunda paylaşılan “Daily Minimum Temperatures in Melbourne” veri setidir. Veri seti, 1981–1990 yılları arasında Melbourne şehrinde ölçülen günlük minimum sıcaklık değerlerini içermektedir. Çalışma kapsamında veri seti üzerinde herhangi bir yapısal değişiklik yapılmamış, yalnızca model eğitimi için gerekli ön işleme adımları uygulanmıştır.

Veri seti bağlantısı:

<https://www.kaggle.com/datasets/paulbrabban/daily-minimum-temperatures-in-melbourne>

Ek C: Model Servisleme ve Kullanıcı Arayüzü

Daily Temperature Prediction (LSTM)

Son 7 günün minimum sıcaklık değerlerini girerek bir sonraki günün tahminini elde edebilirsiniz.

14,15,16,17,18,19,20

Tahmin edilen sıcaklık: 18.18 °C

Clear

Submit

Flag

Examples

10,11,12,13,12,11,12

5,6,7,8,9,10,11

14,15,16,17,18,19,20

Şekil 3 Gradio kullanıcı arayüzü

Şekil 3’te geliştirilen LSTM modelinin Gradio kütüphanesi kullanılarak oluşturulan kullanıcı arayüzü gösterilmektedir. Bu arayüz sayesinde kullanıcılar, son 7 güne ait minimum sıcaklık değerlerini girerek bir sonraki günün tahmin edilen sıcaklık değerini gerçek zamanlı olarak

elde edebilmektedir. Modelin servis edilmesi, alıřmanın uygulamaya ynelik ynn ortaya koymakta ve geliřtirilen sistemin gerek dnya senaryolarında kullanılabilir olduėunu gstermektedir.