T.R.

GEBZE TECHNICAL UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

SOLAR ENERGY PRODUCTION FORECASTING

GÜLNİHAL AKDEM

SUPERVISOR
ASST. PROF. BURCU YILMAZ

GEBZE
2025

T.R.
GEBZE TECHNICAL UNIVERSITY
FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

SOLAR ENERGY PRODUCTION FORECASTING

GÜLNİHAL AKDEM

SUPERVISOR
ASST. PROF. BURCU YILMAZ

2025
GEBZE

This study has been accepted as an Undergraduate Graduation Project in the Department of Computer Engineering on 09/10/2024 by the following jury.

JURY

Member
(Supervisor) : Asst. Prof. Burcu Yılmaz

Member : Assoc. Prof. Habil Kalkan

ABSTRACT

The increasing demand for accurate energy forecasting has led to the development of advanced prediction models, one of which is solar energy.
This study aims to develop a short-term solar energy forecasting model using historical data. Two models were utilized for this purpose: the Encoder-Decoder LSTM with Multivariate inputs, a deep learning approach known for its ability to capture complex temporal patterns, and Facebook's Prophet model, a widely used tool for forecasting time series data.
The relevant datasets were sourced from the EPIAŞ and NASA websites, providing a robust foundation for model training and evaluation. Forecasting models, such as these, have become highly valuable in energy systems, as they assist in improving the efficiency of energy production, reducing costs, and integrating renewable energy sources into the grid.

ÖZET

Son yıllarda enerji tahminine olan talep artmaktadır ve tahmin modellerinin geliştirlmesine yol açmaktadır; bunlardan biri ise güneş enerjisidir.
Bu çalışmada, geçmiş verilere dayalı kısa vadeli güneş enerjisi tahmin modeli geliştirmeyi amaçlamaktadır. Bu amaçla iki model kullanıldı; Birincisi Multivaraite Encoder-Decoder LSTM ve ikinci ise Facebook tarafından geliştirilen Prophet modeli.
Verisetlerini, model eğitimi ve değerlendirilmesi için EPIAŞ ve NASA web sitelerinden temin edilmiştir. Bu tür tahmin modelleri, enerji üretim verimliliğini artırmaya, maliyetleri düşürmeye ve yenilenebilir enerji kaynaklarını enerji şebekesine entegre etmeye yardımcı olduğundan, enerji sistemlerinde son derece değerli hale gelmiştir.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have supported and assisted me throughout the course of this study. First and foremost, I would like to thank my supervisor, Burcu YILMAZ, for their guidance, invaluable advice, and constant encouragement during the entire research process.

I would also like to express my sincere gratitude to Özgün YÜCEL and Habil KALKAN for their valuable insights, continuous support, and assistance in various aspects of this project. Their contributions have greatly enriched this research.

I would also like to acknowledge the EPIAŞ and NASA websites for providing the datasets used in this study.

Finally, I would like to express my deep appreciation to my family and friends for their unwavering support and understanding throughout this journey.

Gülnihal AKDEM

LIST OF SYMBOLS AND ABBREVIATIONS

Symbol or Abbreviation	:	**Explanation**
LSTM	:	Long Short-Term Memory
MinMaxScaler	:	Minimum Maximum Scaler
X	:	the original data,
X_{min}	:	the minimum value of the data
X_{max}	:	the maximum value of the data
X_{norm}	:	the normalized data
GES	:	Güneş Enerji Santrali

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. DATA PREPARATION

The data preparation phase primarily involves data collection, merging of the collected data, transformation, reduction, splitting and normalization. This is a crucial part of the model, as the data must be sufficiently large and meaningful for training and testing the model. Having a sufficiently large and meaningful dataset is critical for making accurate predictions. 1.1.

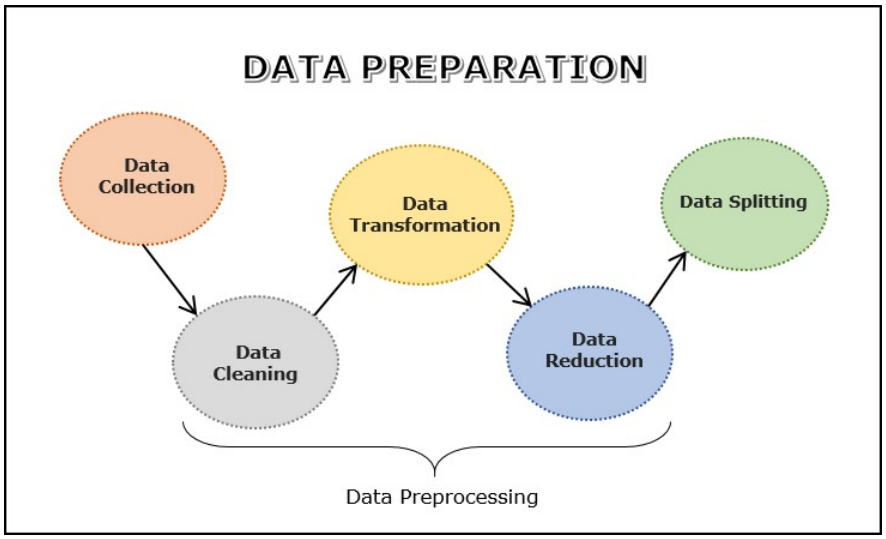

Figure 1.1: Data Preparation Steps

1.1. Data Collection

Data collection is the process of collecting and evaluating information or data from multiple sources to find answers to research problems, answer questions, evaluate outcomes, and forecast trends and probabilities.

In this project, data was collected for forecasting purposes from two primary sources: EPIAŞ1.3 and NASA1.2. Hourly solar energy production data was obtained from EPIAŞ, while satellite data recorded by NASA was used for additional inputs. The data from EPIAŞ corresponds to the target variable, whereas the data from NASA represents the independent variables.

To retrieve data from NASA, several selections must be made, including the coor-

dinates, the required independent variables, and whether the data is available on an hourly or daily basis. Once these parameters are specified, the dataset can be successfully loaded. On the other hand, for EPIAŞ, the dataset can be obtained by selecting the desired date range and solar power plant.

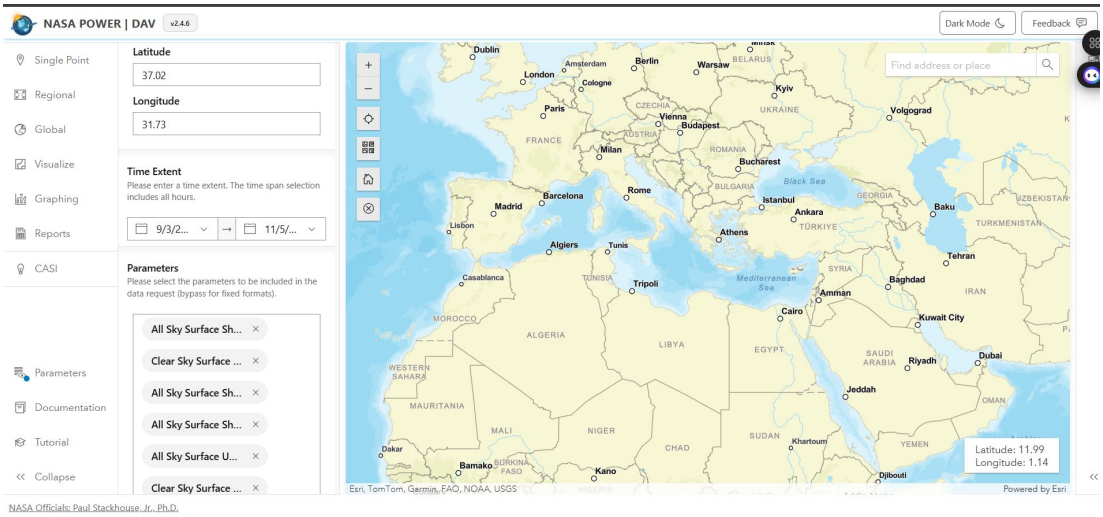

Figure 1.2: NASA

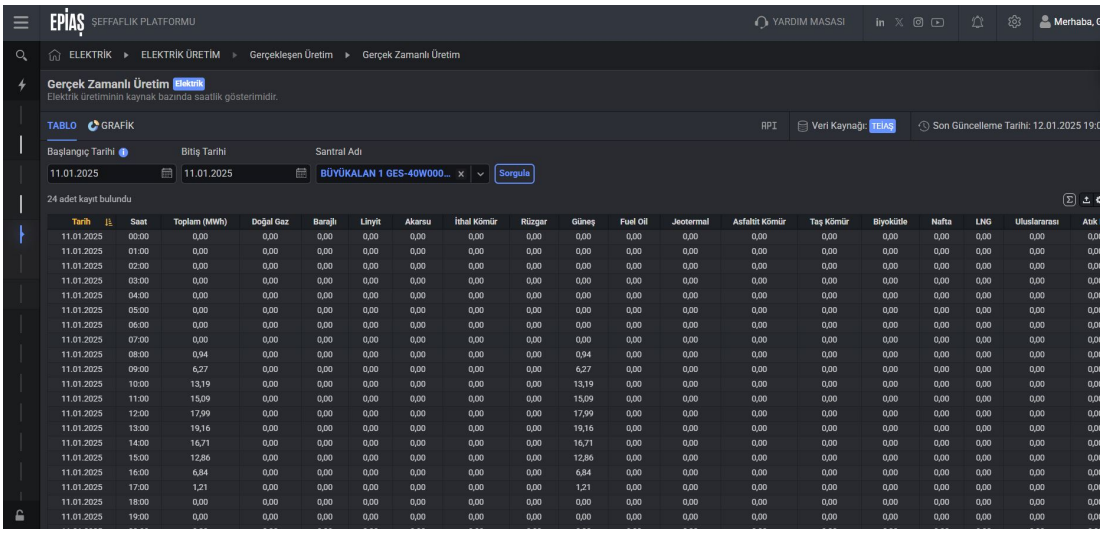

Figure 1.3: EPİAŞ

1.2. Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

Missing values or NaN values were checked. In the NASA dataset, there was a header section. Unnecessary headers were removed. Some columns were merged, and as a result, those rows were deleted. For example, the columns for year, month, day, and hour were initially separate, but after being consolidated into a single ”datetime” column, these individual columns were removed.

1.3. Data Merging

The intersection of the time periods from the data obtained from NASA and EPIAŞ was examined. Only the data that corresponded to the same time periods was used, as the target variables were matched to the independent variables for that specific time frame. This synchronized data was then used for training and teaching the model.

1.4. Data Transformation

Data transformation is the process of converting, cleansing, and structuring data into a usable format that can be analyzed to support decision making processes, and to propel the growth of an organization.[1]

The downloaded data was organized into a DataFrame. Values such as -999.99, which indicated hours with no sunlight, were replaced with 0. In the columns, time was consolidated as a single entity and organized into a separate column.

1.5. Data Normalization

Data normalization is a vital pre-processing, mapping, and scaling method that helps forecasting and prediction models become more accurate. The current data range is transformed into a new, standardized range using this method. Normalization is extremely important when it comes to bringing disparate prediction and forecasting techniques into harmony. Data normalization improves the consistency and comparability of different predictive models by standardizing the range of independent variables or features within a dataset, leading to more steady and dependable results.[2]

Normalisation, which involves reshaping numerical columns to conform to a standard scale, is essential for datasets with different units or magnitudes across different features. Finding a common scale for the data while maintaining the intrinsic variations in value ranges is the main goal of normalization. This usually entails rescaling the features to a standard range, which is typically between 0 and 1. Alternatively, the features can be adjusted to have a mean of 0 and a standard deviation of 1.

In the project, MinMaxScaler was used for normalization.
MinMaxScaler stands for Minimum-Maximum Scaler. It is a method used to transform data into a specific range, usually between 0 and 1.

The purpose of MinMaxScaler is to normalize the data so that all the features have the same scale.

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Figure 1.4: Formula of the MinMaxScaler

X represents the original data,
X_{min} is the minimum value of the data,
X_{max} is the maximum value of the data,
X_{norm} is the normalized data.

2. METHODS

Two models were developed in this project. These are the Encoder-Decoder LSTM with Multivariate and Prophet models.

2.1. Encoder-Decoder LSTM with Multivariate

Multivariate Inputs: LSTMs directly support multiple parallel input sequences for multivariate inputs.

An encoder-decoder LSTM is a model comprised of two sub-models: one called the encoder that reads the input sequences and compresses it to a fixed-length internal representation, and an output model called the decoder that interprets the internal representation and uses it to predict the output sequence.

The project was developed as an LSTM-based multivariate multi-step time series forecasting model for solar power production. Key components include: data splitting into training, validation, and test sets; LSTM model development; prediction; evaluation (metrics: RMSE, MSE, MAE, R^2); and plotting (training and validation loss).

2.2. Prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.[3]

In the Prophet model of this project, the data prepared during the Data Preparation phase is used for model fitting, followed by forecasting. The model is trained using the historical data, and after the fitting process, predictions are made for the future. These predicted values are then visualized by plotting them alongside actual values, providing insights into the forecast's accuracy. To evaluate the model's performance, several metrics are calculated, including RMSE, MSE, MAE, and R^2. Additionally, multivariate input was incorporated into the model to improve the forecasting accuracy by considering multiple variables in the prediction process.

2.3. Metrics, Graphs and User Interface

For two different solar power plants, both the Encoder-Decoder LSTM and Prophet models were trained and tested. In this process, the training-validation loss graphs were plotted, the predicted graphs by the models were visualized, and the performance metrics for the test dataset, including RMSE, MSE, and MAE, were calculated.

During the training process, after 20 epochs, there was a significant sudden increase in validation loss. Therefore, I stopped the learning process at the 20th epoch.

2.3.1. Metrics

Table 2.1: Comparison of model performance metrics for Karapınar GES.

Model	R^2	MAE	RMSE	MSE
Encoder-Decoder LSTM	0.742	2.552	4.107	16.866
Prophet	0.657	1.9116	2.7024	7.3032

Table 2.2: Comparison of model performance metrics for Büyük Alan GES.

Model	R^2	MAE	RMSE	MSE
Encoder-Decoder LSTM	0.603	3.590	5.885	34.627
Prophet	0.709	2.566	3.746	14.0349

2.3.2. User Interface and Graphs

Django is used for displaying the results of the models in the project. Django is an open-source, high-level framework written in Python, used for developing web applications, and it offers built-in features such as rapid development, security, and database management.

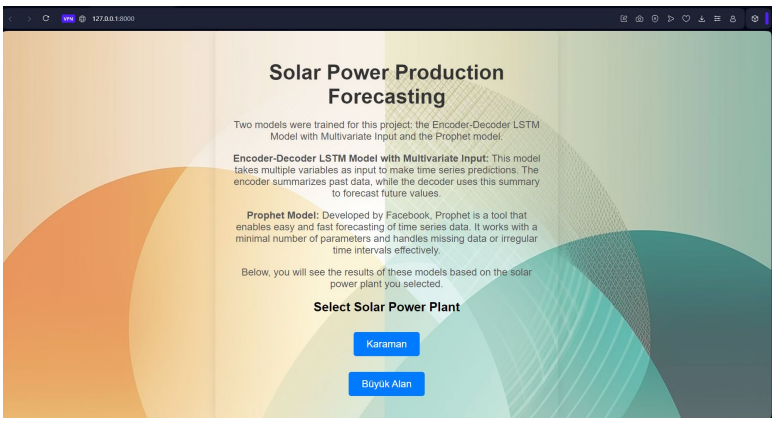

Figure 2.1: Main Page for User Interface

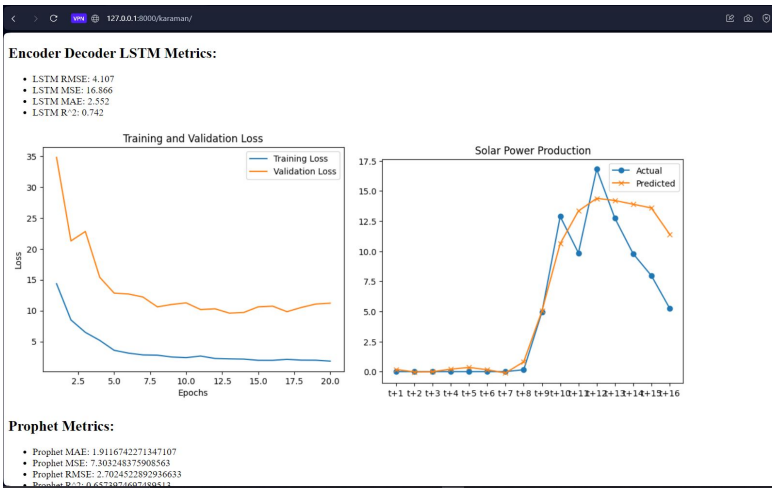

Figure 2.2: Karapınar GES Results

Figure 2.3: Karapınar GES Results

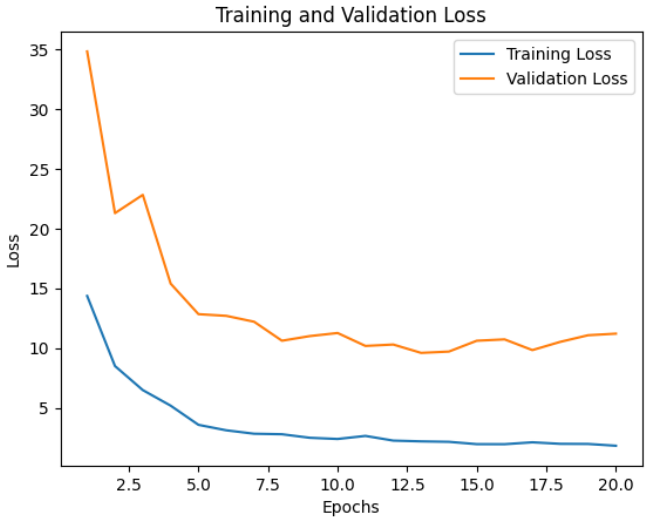

Figure 2.4: Training and Validation Loss Graph for Karapınar GES

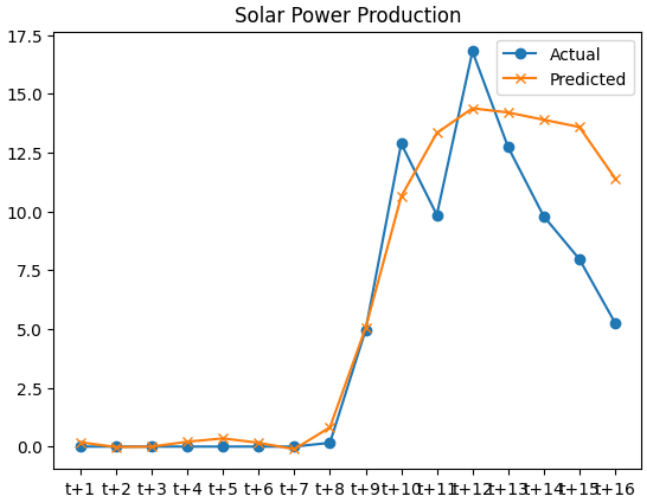

Figure 2.5: The forecasting graph of the Karapınar GES using the LSTM model.

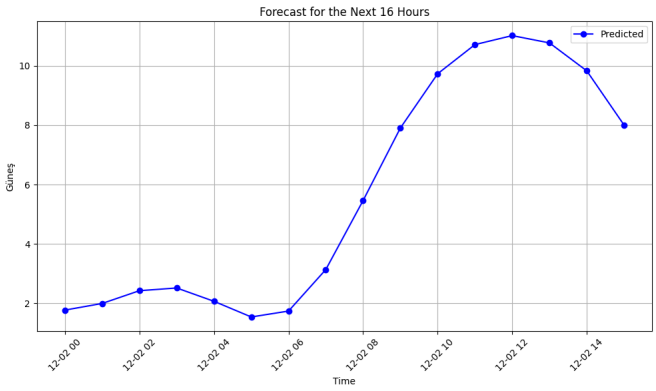

Figure 2.6: The forecasting graph of the Karapınar GES using the Prophet model.

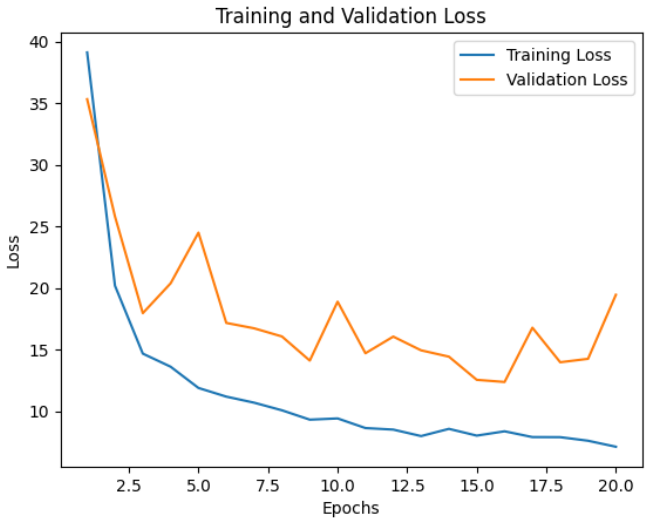

Figure 2.7: Training and Validation Loss Graph for Büyük Alan GES

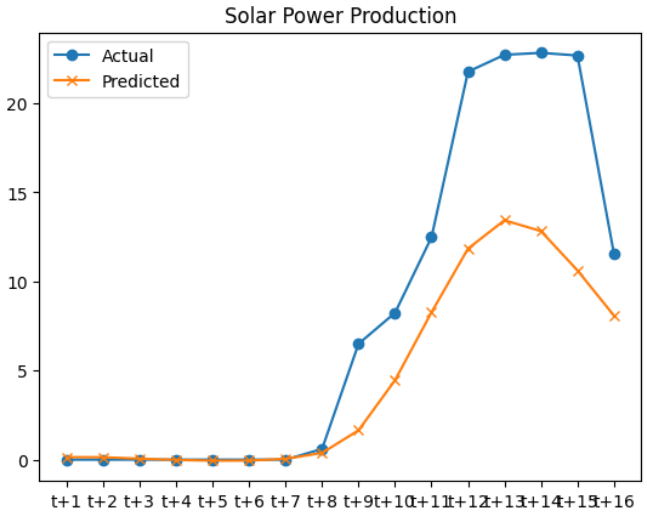

Figure 2.8: The forecasting graph of the Büyük Alan GES using the LSTM model.

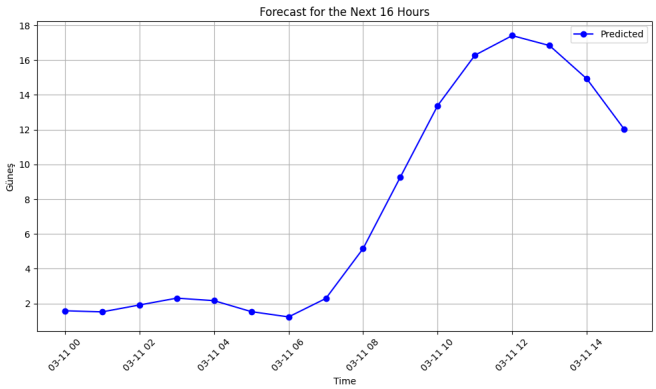

Figure 2.9: The forecasting graph of the Büyük Alan GES using the Prophet model.

3. CONCLUSIONS

The objective of this project is to leverage historical data to develop two distinct models, namely the Encoder-Decoder LSTM and Prophet models, in order to make short-term solar energy production forecasts for two different solar power plants, Karapınar and Büyük Alan. The project aims to observe and analyze the results provided by these models in predicting solar energy output.

BIBLIOGRAPHY

[1] T. S. Inc., *What is data transformation?* Accessed: 2025-01-12, 2023. [Online]. Available: https://www.tibco.com/glossary/what-is-data-transformation#:~:text=Data%20transformation%20is%20the%20process,that%20of%20the%20destination%20system.

[2] GeeksforGeeks, *What is data normalization?* Accessed: 2025-01-12, 2023. [Online]. Available: https://www.geeksforgeeks.org/what-is-data-normalization/.

[3] Facebook, *Prophet: Forecasting at scale*, Accessed: 2025-01-12, 2023. [Online]. Available: https://facebook.github.io/prophet/.

[4] NASA Langley Research Center, *Nasa power data access viewer*, Accessed: 2025-01-12, 2023. [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/?edit=true&appid=db2adab099f246f6b4d4c113f2a5ea09.

[5] T. Software, *What is data cleaning?* Accessed: 2025-01-12, 2023. [Online]. Available: https://www.tableau.com/learn/articles/what-is-data-cleaning#:~:text=Data%20cleaning%20is%20the%20process,to%20be%20duplicated%20or%20mislabeled.

[6] EPIAŞ, *Epiaş - enerji piyasaları işletme a.ş.* Accessed: 2025-01-12, 2023. [Online]. Available: https://www.epias.com.tr/.

[7] PNGTree, *Geometric simple background for website*, Accessed: 2025-01-12, 2023. [Online]. Available: https://tr.pngtree.com/freebackground/geometric-simple-background-for-website_2443633.html.

[2] [4] [1] [5] [6] [3] [7]

CV

europass

Gülnihal Akdem

Date of birth: 24/12/1997 | **Nationality:** Turkish | **Email address:** gulnihalakdem1@gmail.com |
Email address: gulnihalakdem@gmail.com | **Website:** https://github.com/akdemgulnihal |
LinkedIn: https://www.linkedin.com/in/gülnihal-akdem-35a154153/?trk=public-profile-join-page

ABOUT ME

Fourth-year Computer Engineering student. I am looking for an internship where I can develop myself and gain practical experience. Moreover, You can see some projects on my Github profile.

EDUCATION AND TRAINING

01/09/2019 – CURRENT
COMPUTER ENGINEERING Gebze Technical University

18/07/2022 – 12/08/2022
INTERNSHIP Faturamatik Elektronik Para ve Ödeme Kuruluşu A.Ş.

01/09/2017 – 01/06/2019
ENGLISH LANGUAGE PREPARATORY PROGRAM İhsan Doğramacı Bilkent University

LANGUAGE SKILLS

Mother tongue(s): **TURKISH**

Other language(s): **ENGLISH (IELTS: 6.0 IN 2018)**

DIGITAL SKILLS

Python | C | C++ | FPGA programming with Verilog HDL | Linux | Windows | RDBMS (Relational Database Management Systems) | Java | MySQL | MsSQL | Html, Java, CSS | Javascript(Nodejs, Expressjs) | Qt Framework | MongoDB | Basic Agile & Scrum knowledge | MIPS Assembly | Git(Github) | Microsoft Office (Outlook, Excel, Word, PowerPoint) | LaTex: used for documents preparation (e.g. reports and presentations). | Common Lisp

DRIVING LICENCE

Driving Licence: B1

Driving Licence: B

PROJECTS

Graduation Project I

Turkish Text to SQL Query Conversion using Deep Learning (2023-2024 Spring)

Software Project - SportKids (Scrum Master & Developer)

The web part of the project used Node.js, while the mobile part was developed with Kotlin. Firebase's Firestore database was used for data storage. I worked as a developer on the web side of the project and also served as the Scrum Master.

The purpose of the project is to enable parents to monitor their children's progress at sports school. Parents can log in via the mobile app. On the web side, functionalities related to children, parents, and courses are managed.

CURRENT
Graduation Project II

Solar Power Production Forecasting using Deep Learning (2024-2025 Fall)