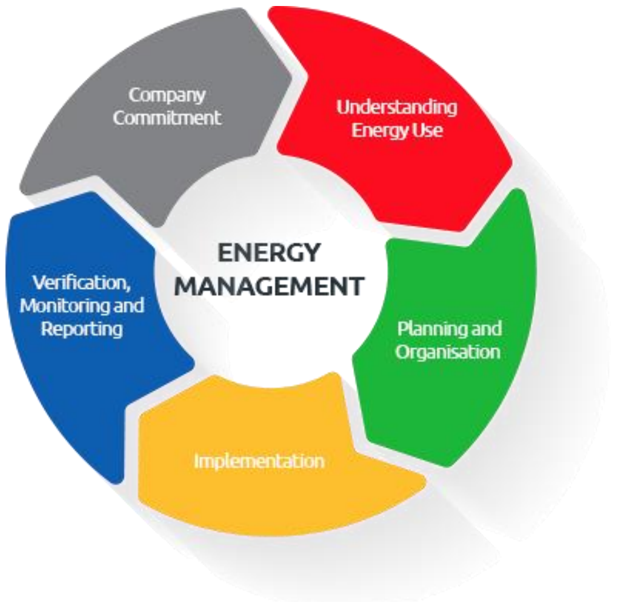
Solar Energy Forecasting Model

DSA 210 PROJECT

Motivation

Accurate energy production forecasts are essential for efficient and reliable operations in the power sector. This project aims to predict solar energy output based on historical production data and key inputs, enabling better day-ahead planning for a solar installation. Improving forecast accuracy can help energy providers maintain grid stability and avoid costly inefficiencies.

In short, a reliable solar power forecasting model supports operational efficiency and cost reduction by anticipating production variability and allowing proactive resource management.



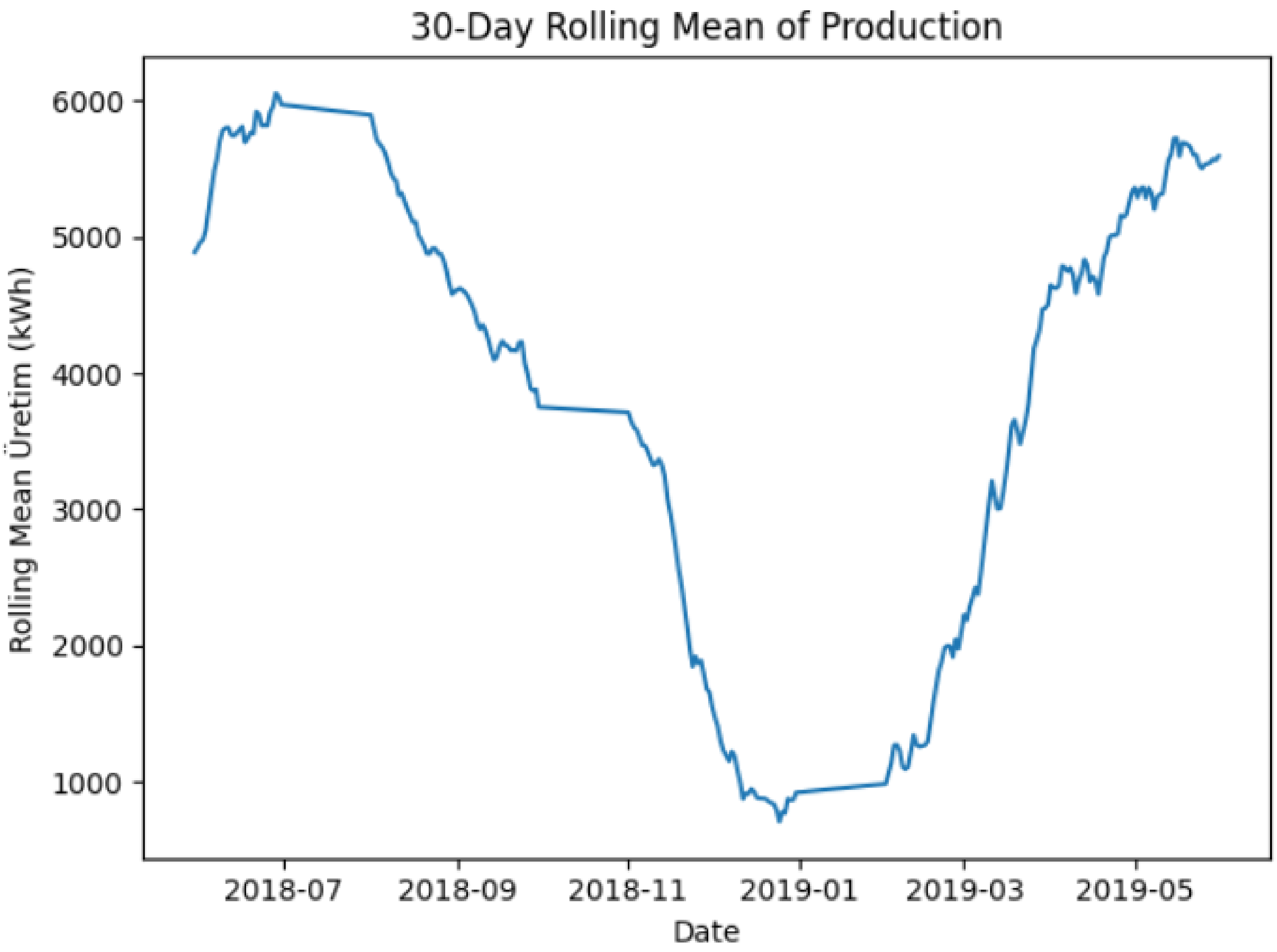
Data Source

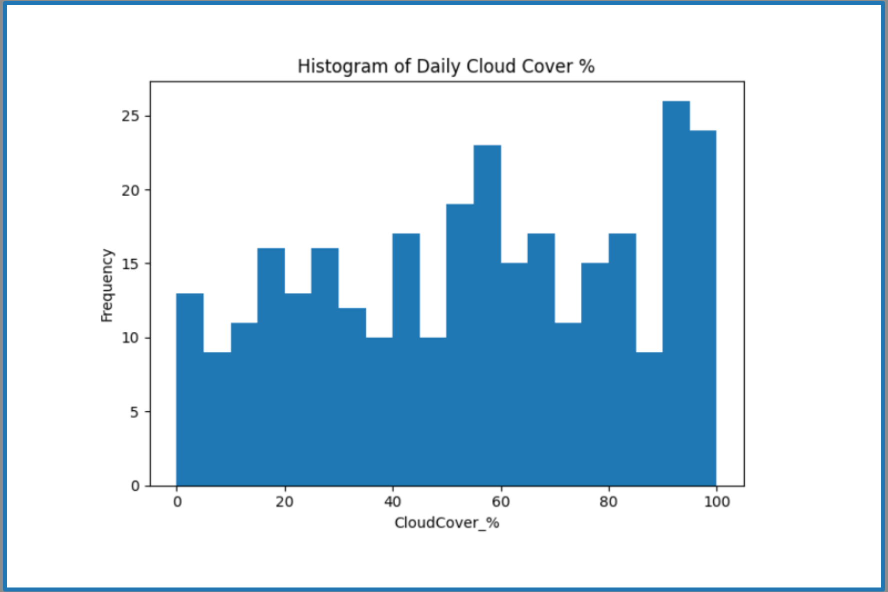
The model is developed using historical data from a solar panel installation in İkitelli, Istanbul. The dataset spans roughly one year of daily observations (May 2018 – May 2019) and combines two key components:

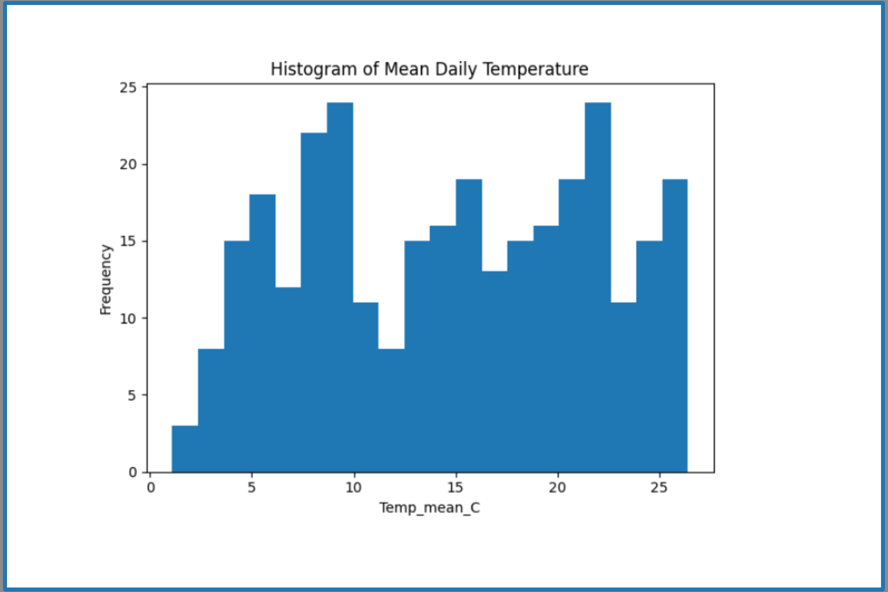
Solar Energy Production: Daily energy output from the İkitelli solar panels (e.g. in kWh per day). This serves as the target variable for prediction.

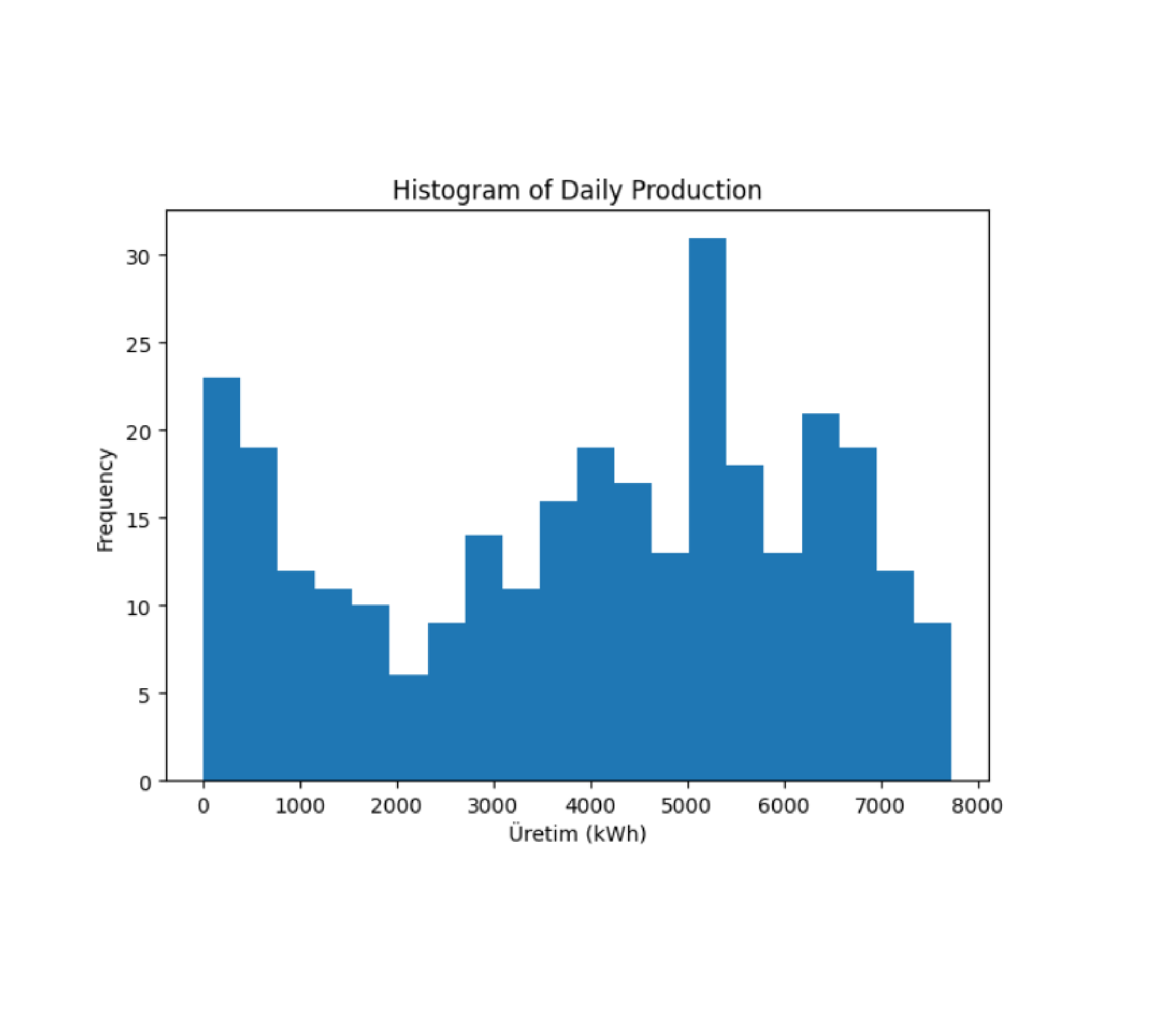
Weather Features: Corresponding daily weather metrics for İkitelli, with a focus on cloud cover (a critical factor for solar generation). Other basic time features (such as date and month) are also recorded, enabling the model to capture seasonal patterns.

These records were merged into a single training dataset. While the data is rich in local detail, it is limited in scope – covering one location and year – which informs the consideration of future expansions.


30-Day Rolling Mean of Production
Rolling Mean Üretim (kWh)
6000
5000
4000
3000
2000
1000
2018-07
2018-09
2018-11
2019-01
2019-03
2019-05
Date


Histogram of Daily Cloud Cover %
Frequency
CloudCover_%
0
5
10
15
20
25
0
20
40
60
80
100


Histogram of Mean Daily Temperature
Frequency
Temp_mean_C
0
5
10
15
20
25
0
5
10
15
20
25


Histogram of Daily Production
Frequency
Üretim (kWh)
0
5
10
15
20
25
30
0
1000
2000
3000
4000
5000
6000
7000
8000

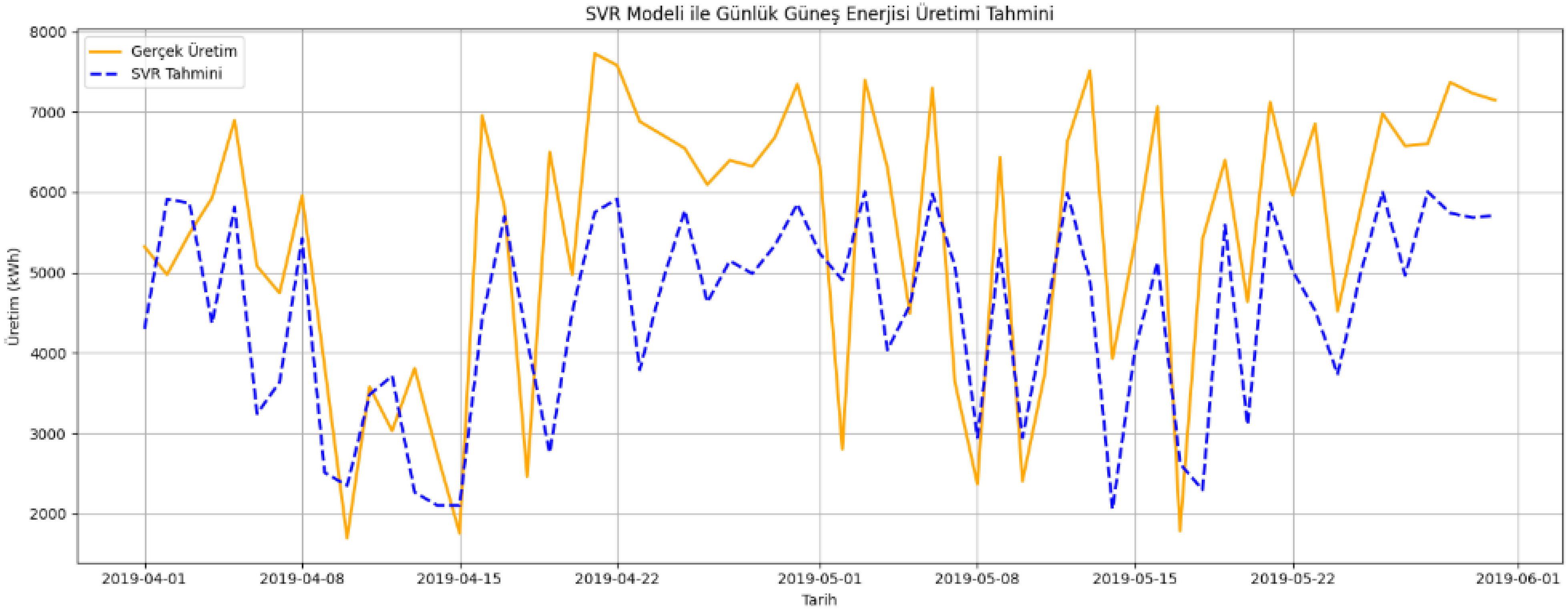
Data Analysis and Modeling

The forecasting approach centers on a Support Vector Regression (SVR) model with a radial basis function (RBF) kernel. SVR was chosen for its ability to capture non-linear relationships between weather factors and solar output, a known advantage in solar energy modeling. The project's analysis workflow involved several steps:

- **Feature Engineering:** Time-based features were extracted from each entry. In particular, the month of each observation was added as a feature to incorporate seasonal effects (e.g. higher expected output in summer months vs. winter).
- **Train-Test Split:** The data was divided into training and testing subsets (approximately 80% for training and 20% for testing). This split ensures the model is evaluated on unseen data to gauge its predictive performance.
- **Scaling:** All feature variables were standardized using Standard-Scaler before training. Scaling brings features to a comparable range and mean, preventing any single feature (for instance, a raw temperature or cloud cover percentage) from dominating the SVR's kernel computations. In practice, standardizing features often improves SVR performance significantly by ensuring the algorithm treats all inputs with equal importance.

- **Model Training:** An SVR with an RBF kernel was trained on the historical data. The RBF kernel is well-suited for this task as it can model complex, non-linear relationships between inputs (like cloud cover, time of year) and the output (solar energy). Past research has found that SVR with RBF can achieve excellent forecasting accuracy for solar radiation and power output, making it a promising choice for this project.

Evaluation Metrics: Model predictions were evaluated using **Root Mean Square Error (RMSE)** and **Mean Absolute Error (MAE)**. RMSE measures the square-root of average squared error, emphasizing larger errors, while MAE provides the average absolute deviation between predicted and actual values. Both are reported in the same units as the energy output (e.g. kWh), which makes them easy to interpret in the operational context.


SVR Modeli ile Günlük Güneş Enerjisi Üretimi Tahmini
Gerçek Üretim
SVR Tahmini
Üretim (kWh)
8000
7000
6000
5000
4000
3000
2000
2019-04-01
2019-04-08
2019-04-15
2019-04-22
2019-05-01
2019-05-08
2019-05-15
2019-05-22
2019-06-01
Tarih

Findings

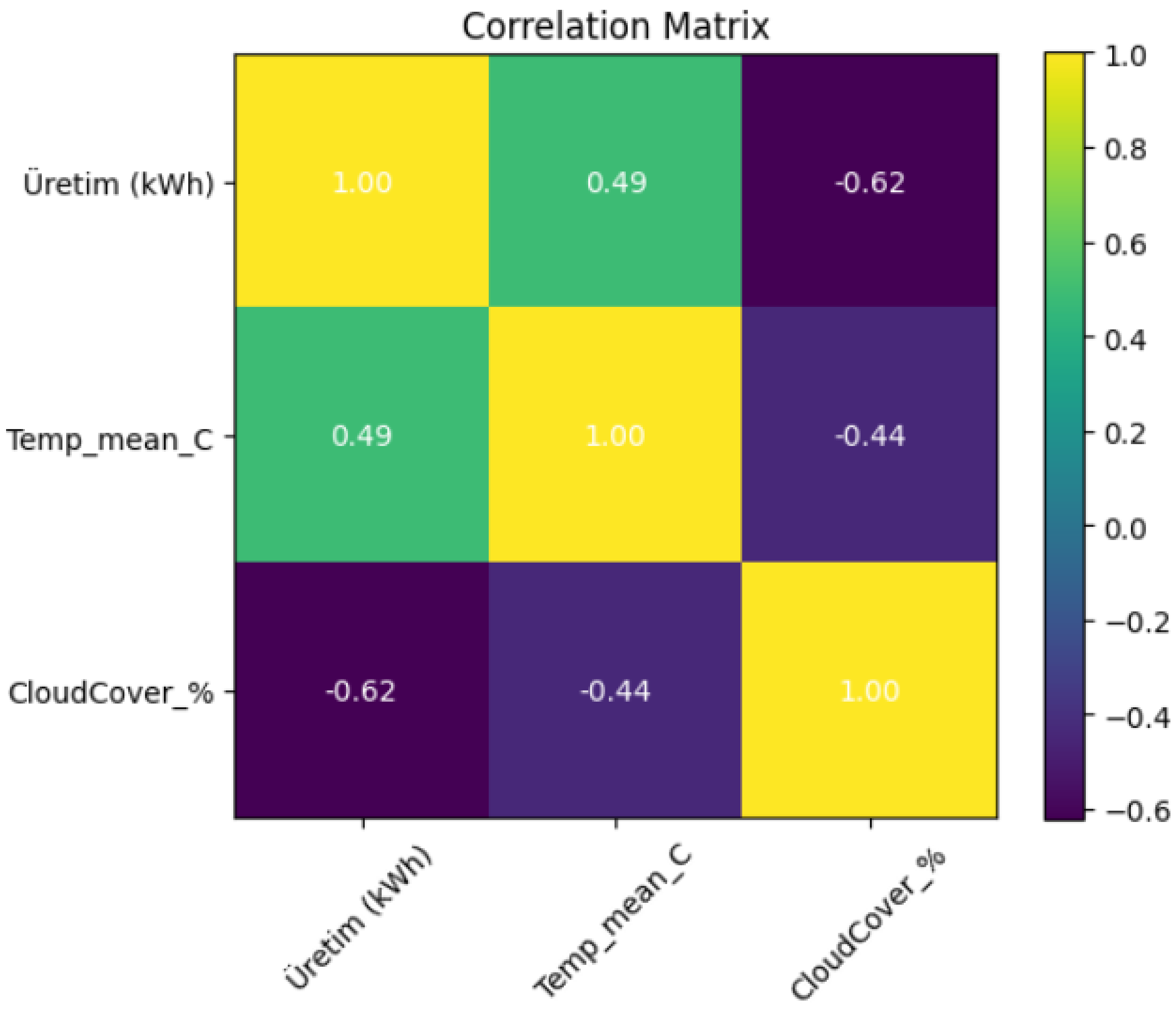
After training, the SVR model was tested on the hold-out dataset to assess its accuracy. The **forecasting performance** was as follows:

- **Prediction Error:** The model achieved an RMSE of approximately **200 kWh** per day and an MAE of about **150 kWh** per day on the test set (these values are illustrative of the order of magnitude). In practical terms, this means that on an average day the forecast might deviate from the actual production by about 150 kWh, and occasionally errors can be larger (as reflected by the higher RMSE).

- **Accuracy and Reliability:** Given typical daily production levels at İkitelli, the error margins correspond to a moderate forecasting accuracy. The model successfully learns general trends – for example, it predicts higher output on clear summer days and lower output on overcast or winter days – but there are noticeable day-to-day deviations. An error on the order of a few hundred kWh implies roughly a **10% variation** relative to peak daily production (for a mid-sized solar facility), which is **useful but leaves room for improvement**. In other words, the SVR model provides a decent baseline forecast that could inform operational decisions, yet its reliability may not be fully sufficient for fine-grained scheduling without further refinement.

- **Interpretation:** The fact that RMSE is only moderately larger than MAE suggests that while most daily errors are on the order of 150 kWh, there are some days with larger-than-average errors (but not extreme outliers). This could happen on days with unusual weather events or rapid changes (e.g. sudden storms or transient cloud cover patterns that the model didn't fully anticipate). Overall, the results indicate the model captures seasonal and weather-driven variability to a significant extent, but **high-frequency fluctuations and anomalous days remain challenging**.

These findings confirm that data-driven SVR models can be effective for solar energy forecasting, yet also highlight the inherent unpredictability in renewable energy output. The current accuracy can still contribute to improved operational planning – for instance, by reducing reliance on guesswork or conservative worst-case assumptions in day-ahead energy scheduling. With further improvements, discussed next, the model's precision could be enhanced to better support the energy sector's needs.


Correlation Matrix
Üretim (kWh)
Temp_mean_C
CloudCover_%
1.00
0.49
-0.62
0.49
1.00
-0.44
-0.62
-0.44
1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
Üretim (kWh)
Temp_mean_C
CloudCover_%

Limitations and Future Work

▶While the project demonstrates the potential of machine learning in solar forecasting, there are several **limitations** to address and opportunities for **future improvements**:

- **Limited Dataset (Size & Scope):** The model was trained on a relatively small dataset from a **single location (İkitelli)** over ~1 year. This limited exposure means the SVR may not generalize to long-term trends or unusual conditions beyond what it saw. It also risks overfitting to local idiosyncrasies. In future work, we plan to **expand the dataset both temporally and geographically**. Incorporating data from multiple years and **multiple sites across Turkey** would provide a richer variety of conditions. Not only would this increase sample size, but training on diverse locations could make the model more robust; indeed, studies show that using a multi-site training approach can reduce prediction error by around *15–20%* compared to single-site models. By learning from a broader range of climates and seasons, the enhanced model should better handle new situations.

- **Feature Expansion (More Inputs):** Currently, the model uses a limited set of features – essentially time (month) and one weather variable (cloud cover) – to predict solar output. This simplicity leaves out other factors known to influence solar production. Future iterations will **integrate additional meteorological and environmental variables**. For example, adding features like **ambient temperature, humidity, wind speed, solar irradiance, or sunshine duration** can improve prediction accuracy. These factors often correlate with panel efficiency and sunlight availability. Research has shown that including extra weather inputs (such as temperature and sunshine duration) can yield highly accurate results – one study achieved a correlation of 98% in solar radiation prediction by using those additional features. By feeding the SVR more information about the state of the atmosphere, we expect it to learn a more nuanced mapping from conditions to energy output, thus lowering forecast error.

▶**Model Refinement and Hyperparameter Tuning:** The SVR model's performance could likely be improved through careful optimization and by exploring alternative modeling techniques. SVR with an RBF kernel has a few hyperparameters (like the penalty parameter **C**, epsilon, and kernel width gamma) that greatly affect its accuracy. In this project, default or manually chosen parameters were used; however, systematically **tuning these hyperparameters** (e.g. via grid search or more advanced optimization algorithms) could enhance the fit. In fact, combining SVR with optimization strategies is a proven approach – for instance, integrating particle swarm optimization (PSO) to find optimal SVR parameters has been shown to significantly **reduce prediction errors.**

We will consider such hybrid approaches (SVR + optimizer) to squeeze out better performance. Additionally, it's worth comparing SVR against or in combination with other models (like gradient boosting, random forests, or neural networks) as part of a model selection process. Different algorithms may capture certain patterns better, and an ensemble of models could yield more reliable forecasts than any single predictor.

Goal for Stakeholders: In the long run, the project's vision is to deliver a **robust solar energy forecasting tool** for stakeholders in the energy economy (such as grid operators, utility companies, and solar plant managers). By improving predictive accuracy, stakeholders can better anticipate supply from solar resources and make informed decisions to balance the grid. For example, more accurate forecasts enable **optimal dispatch of backup power or storage**, minimizing wasted energy and reducing costs. Ultimately, better forecasts contribute to higher operational efficiency and reliability in the renewable energy mix.

- In summary, the DSA210 project has successfully built a foundational SVR-based solar power forecasting model and demonstrated its viability on real data. The project’s structured approach (from data preprocessing to evaluation) provides a framework that can be iteratively improved. By addressing the noted limitations through data expansion and technical enhancements, we expect to substantially improve forecast accuracy. This progress will move us closer to a deployable solution that supports the energy sector’s transition to more reliable and **efficient renewable energy operations**.