

Contents

Abstract.....	2
Introduction.....	3
Literature Survey.....	4
Objective	5
Methodology	5
Machine Learning Techniques	5
How does machine learning work?	5
Supervised learning	6
Unsupervised Learning	6
Prediction of the Flow of Solar Radiation	7
Data Gathering for Solar Energy Prediction.....	8
Data Cleaning.....	9
Regression Model	10
GitHub	12
Flask	14
Docker	15
Conclusion.....	16
References.....	17

ABSTRACT

The increased competitiveness of solar PV panels as a renewable energy source has increased the number of PV panel installations in recent years. In the meantime, higher availability of data and computational power have enabled machine learning algorithms to perform improved predictions. As the need to predict solar PV energy output is essential for many actors in the energy industry, machine learning and time series models can be employed towards this end. In this study, a comparison of different machine learning technique.

Introduction

The global shift towards renewable energy sources (RES) has driven the development of photovoltaic (PV) panels. For example, the costs of producing electricity from PV panels have dropped significantly, while simultaneously increasing the energy conversion efficiency. The decreased cost and increased efficiency have made PV panels a competitive alternative as a RES in many countries. However, since PV panel energy output depends on weather conditions such as cloud cover and solar irradiance, the energy output of the PV panels is unstable. To understand and manage the output variability is of interest for several actors in the energy market. One evident nuisance is the inherited variation of weather, which makes accurate weather forecasting challenging. Parallel to the increased demand of PV power forecasting solutions, the means for forecasting with the help of machine learning (ML) techniques have in recent years gained in popularity relative to traditional time series predictive models. Although ML techniques are nothing new, the improved computational capacity and the higher availability of quality data have made the techniques useful for forecasting. This poses for an interesting area of research when forecasting the solar power output:



Figure 1. Solar Power plant

Historical satellite data, environment data, and real-time weather can all be integrated and analysed by predictive algorithms, in order to inform hardware maintenance decisions. It is not isolated and clearly shows how investing in software and custom machine learning and analytics services can have an impact on the performance and maintenance of expensive hardware (such as solar plants or power networks). By leveraging forecasted energy production information, grid operators can cut down on operational costs and make informed decisions, based on real-time performance data.

Literature Survey

1. Aneela Zameer, Farah Shahid², Mudasser Afzal, Muhammad Hassan, "Intelligent forecast models for daily solar energy prediction", Research gate- October 2020

The above paper shows the exploration of solar energy characteristics and comparison of linear and non-linear machine learning methodologies to improve their generalization ability for better adaption and reduced prediction error. For this purpose, various machine learning based regression models such as multivariate linear regression, ridge regression, lasso regression are implemented for daily solar energy forecast

2. K. Anuradha¹, Deekshitha Erlapally , G. Karuna , V. Srilakshmi , K. Adilakshmi Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques, E3S Web of Conferences 309, 01163 ICMED 2021

The above paper concentrates on solar power generation using photovoltaic (PV) systems all over the world. Because the output power of PV systems is alternating and highly dependent on environmental circumstances, solar power sources are unpredictable in nature. The impacts of various environmental conditions on the output of a PV system are discussed. Machine Learning (ML) algorithms have shown great results in time series forecasting and so can be used to anticipate power with weather conditions as model inputs. The use of multiple machine learning, Deep learning and artificial neural network techniques to perform solar power forecasting.

3. Bouchaib Zazoum Solar photovoltaic power prediction using different machine learning methods, 2021 8th International Conference on Power and Energy Systems Engineering (CPESE 2021)

The above paper concentrates on the problem of automatically generating models that accurately predict renewable generation based on National Weather Service forecasts (NWS). Using historical NWS forecast data and data generated by solar panels, we experiment with a variety of machine learning techniques to develop prediction models. For developing prediction models, a variety of regression algorithms are tested, including linear least squares and support vector machines with various kernel functions. In these tests it shows that a machine learning approach can correctly anticipate solar power radiations.

4. C. Vennila, Anita Titus, T. Sri Sudha, U. Sreenivasulu, N. Pandu Ranga Reddy, K. Jamal, Dayadi Lakshmaiah, P. Jagadeesh and Assefa Belay Forecasting Solar Energy Production Using Machine Learning, Hindawi International Journal of Photoenergy Volume 2022, Article ID 7797488, 7 pages <https://doi.org/10.1155/2022/7797488>

The above paper concentrates on large range of datasets and time steps, prediction ranges, settings, and performance measurements, a single machine learning model cannot improve forecasting performance on a single dataset or time step. With the use of hybrid machine learning algorithms, projections for solar power have also been enhanced. In order to effectively predict the availability of solar power radiations, significant attention has lately been drawn to support vector machines (SVMs) and deep learning algorithms.

Objective

- 1) Data collection of a 450kW solar PV plant
- 2) Perform Data Wrangling on the collected data set.
- 3) Perform Exploratory Data Analysis
- 4) Predict the output generation using appropriate Machine Learning Techniques.
- 5) Develop CI-CD Pipeline for deployment.
- 6) Develop front end and backend setup for actual verification.

Methodology:

Machine Learning Techniques

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to directly "learn" from data without relying on a predetermined equation as a model.

How does machine learning work?

Machine learning uses two techniques: supervised learning, which trains a model on known input and output data to predict future outputs, and unsupervised learning, which uses hidden patterns or internal structures in the input data.

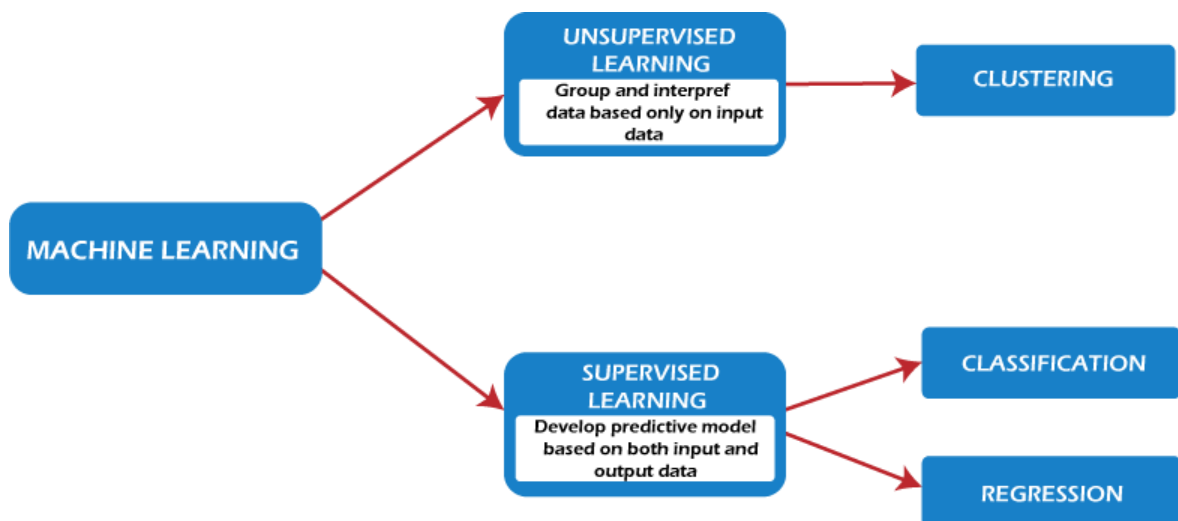


Figure 2. Machine learning types

Supervised learning

Supervised machine learning creates a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to the new data. Use supervised learning if you have known data for the output you are trying to estimate. Supervised learning uses classification and regression techniques to develop machine learning models.

Classification models classify the input data. Classification techniques predict discrete responses. Common algorithms for performing classification include support vector machines (SVMs), boosted and bagged decision trees, k -nearest neighbors, Naive Bayes, discriminant analysis, logistic regression, and neural networks.

Unsupervised Learning

Detects hidden patterns or internal structures in unsupervised learning data. It is used to eliminate datasets containing input data without labeled responses. Clustering is a common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns and clusters in the data. Applications for cluster analysis include gene sequence analysis, market research, and commodity identification.



Figure 3. Clustering patterns

Ten methods are described and it is a foundation you can build on to improve your machine learning knowledge and skills:

- Regression
- Classification
- Clustering
- Dimensionality Reduction
- Ensemble Methods
- Neural Nets and Deep Learning
- Transfer Learning

- Reinforcement Learning
- Natural Language Processing
- Word Embedding's

Let's differentiate between two general categories of machine learning: supervised and unsupervised. We apply supervised ML techniques when we have a piece of data that we want to predict or interpret. We use the previous and output data to predict the output based on the new input.

Prediction of the Flow of Solar Radiation

Our project consisted of three parts data preprocessing, model building, and model prediction. The data preprocessing involved four steps: data quality control, dataset partitioning, data scaling, and variable selection. Among them, data quality control, dataset partitioning, and data scaling are described in Section “Study Area and Datasets,” and variable selection is described in Section

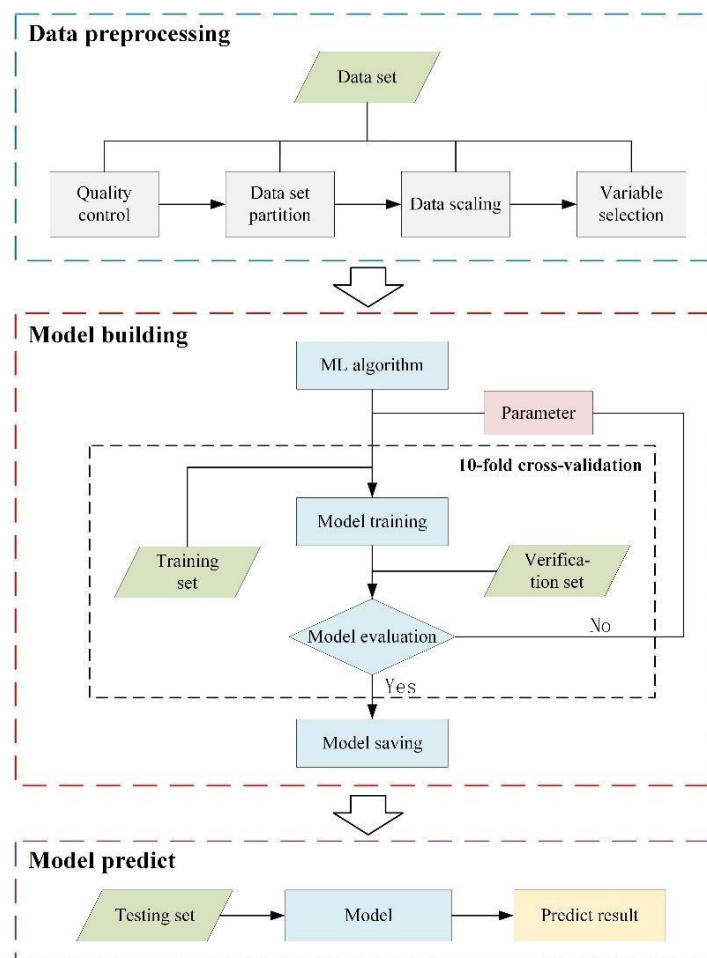


Figure 4. Flow chart of the machine learning models used to estimate solar radiation.

1. Data collection and data preprocessing;
2. Choose a machine learning algorithm from the 12 algorithms to predict solarradiation;
3. Compare solar radiation predictive ability based on different parameters;
4. If the best predictive ability is achieved, save the model;
5. Return to step (2) and choose another machine learning algorithm until all 12algorithms have been subjected to machine learning model building;
6. Input the preprocessing dataset (we prepared datasets on two timescales—daily andmonthly—to estimate the solar radiation predictive performance of the 12 machine learning models) and use the 12 saved machine learning models to predict solar radiation and obtain the predicted results;
7. Save predicted results and analyze.

Data Gathering for Solar Energy Prediction

In order to design for solar energy prediction, it is important to analyze the main factors thataffect solar energy generation. These factors will have to be considered by the model as inputs, which will need to be adequately tailored for each particular situation we wish to predict.

Aside from the energy carried by solar rays, the amount of electric solar energy generated also varies depending on various parameters and conditions that vary depending on the moment of the day. Among these factors, the most notable ones are:

- Solar azimuth angle
- Air temperature
- Wind speed
- Performance of the photovoltaic (PV) panel
- Size of the generating installation

In order to provide a better overview of all factors involved in energy generation, we present an overview

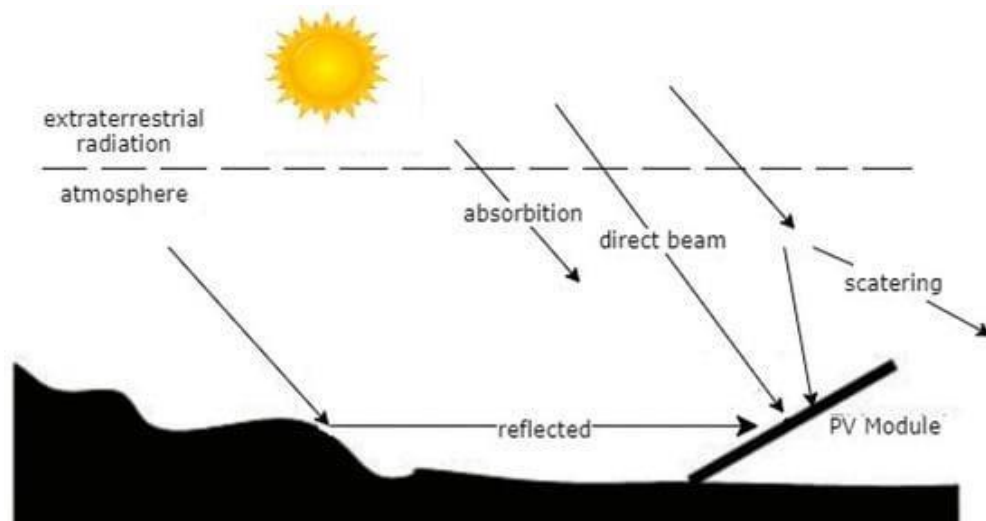


Figure 5. Solar radiation in Earth's atmosphere.

Once we have summarized the factors that affect solar energy generation, in the next section we present how we make use of these factors to predict solar energy output taking into account

the characteristics of the solar panel installation as well as its location and other particular conditions.

Hence the data can be collected using the weather station along with the weather monitoring system

Data Cleaning

Data cleaning is one of the important parts of machine learning. It plays a significant part in building a model. If we have a well-cleaned dataset, there are chances that we can get achieve good results with simple algorithms also, which can prove very beneficial at times especially in terms of computation when the dataset size is large. Obviously, different types of data will require different types of cleaning. However, this systematic approach can always serve as a good starting point.

Steps involved in Data Cleaning:

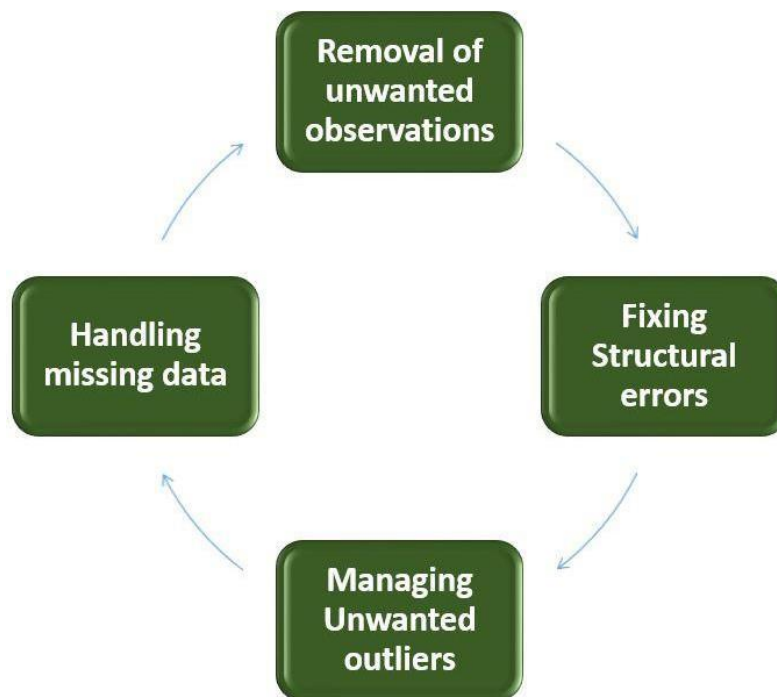


Figure 6. Data cleaning process

1) Removal of unwanted observations

This includes deleting duplicate/ redundant or irrelevant values from your dataset. Redundant observations alter the efficiency by a great extent as the data repeats and may add towards the correct side or towards the incorrect side, thereby producing unfaithful results. Irrelevant observations are any type of data that is of no use to us and can be removed directly.

2) Fixing Structural errors

The errors that arise during measurement, transfer of data, or other similar situations are called structural errors. Structural errors include typos in the name of features, the same attribute with a different name, mislabeled classes, i.e. separate classes that should really be the same, or inconsistent capitalization

3) Managing unwanted outliers

Outliers can cause problems with certain types of models. Sometimes, removing them improves performance, sometimes not. So, one must have a good reason to remove the outlier, such as suspicious measurements that are unlikely to be part of real data.

4) Handling missing data

Missing data is a deceptively tricky issue in machine learning. We cannot just ignore or remove the missing observation. They must be handled carefully as they can be an indication of something important. The two most common ways to deal with missing data are:

- a) Dropping observations with missing values.
- b) Imputing the missing values from past observations.

So, missing data is always an informative and an indication of something important. And we must be aware of our algorithm of missing data by flagging it. By using this technique of flagging and filling, you are essentially allowing the algorithm to estimate the optimal constant for messiness, instead of just filling it in with the mean.

Regression Model

Regression methods fall under the category of supervised ML. They help predict or interpret a particular numerical value based on prior data, such as predicting an asset's price based on past pricing data for similar properties. The simplest method is linear regression, where we use the mathematical equation of the line ($y = m * x + b$) to model the data set. We train a linear regression model with multiple data pairs (x, y) by computing the position and slope of a line that minimizes the total distance between all data points and the line. In other words, we calculate the slope (M) and the y-intercept (B) for a line that best approximates the observations in the data.

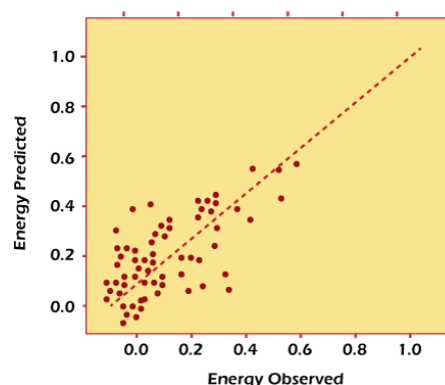


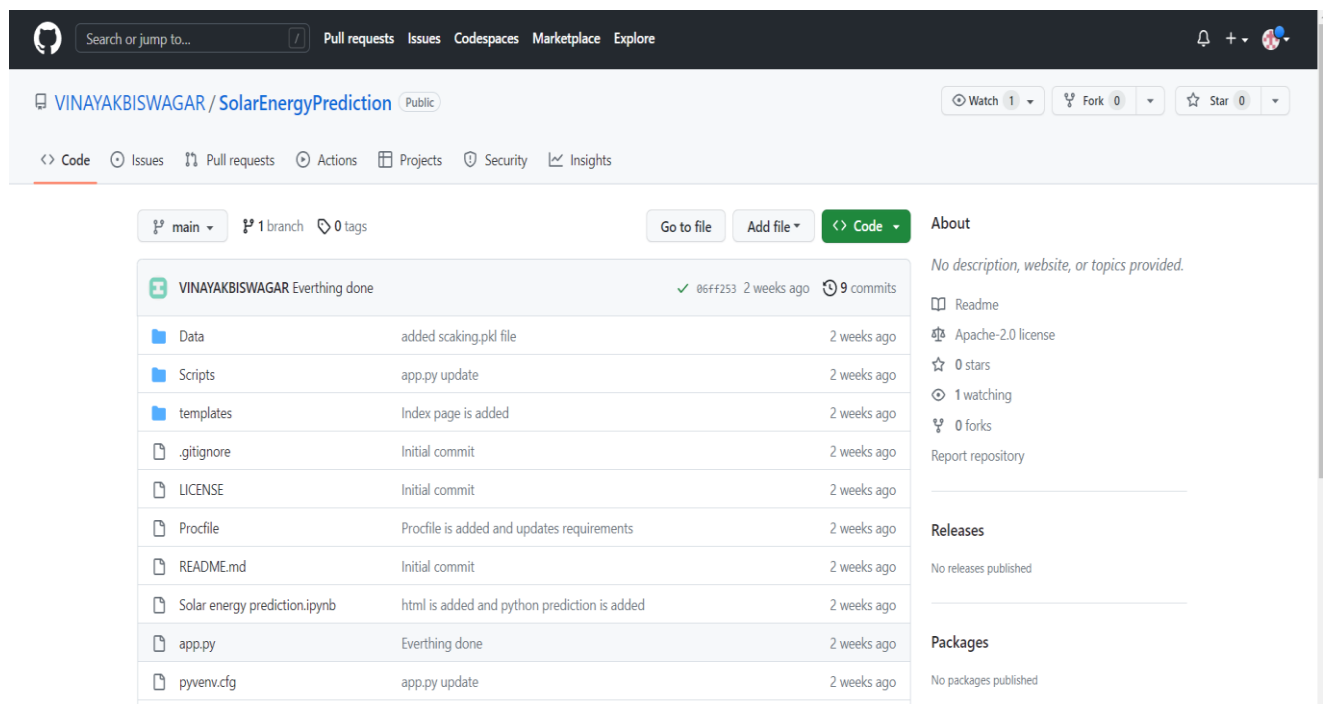
Figure 7. Linear regression model estimates of building energy consumption (kWh).

Regression techniques run the gamut from simple (linear regression) to complex (regular linear regression, polynomial regression, decision trees, random forest regression, and neural nets). But don't get confused: start by studying simple linear regression, master the techniques, and move on.

GitHub:-

GitHub is one of the largest online communities of developers, with over 56 million registered users and more than 100 million code repositories. It has become a central hub for open source software development, where developers from all around the world can share their code and collaborate with others.

One of the most significant aspects of GitHub is its role in promoting open source software development. The platform has made it much easier for developers to share their code with others and collaborate on open source projects. As a result, GitHub has become an essential tool for many organizations and individuals who are passionate about open source software.



<https://github.com/VINAYAKBISWAGAR/SolarEnergyPrediction>

Here are some of the key features of GitHub:

- **Code Hosting:** GitHub provides a place for developers to host their code repositories, making it easy to share their code with others and collaborate on projects.
- **Version Control:** GitHub uses Git, a popular version control system that allows developers to keep track of changes to their code over time. This means that GitHub can keep track of all the changes made to a project, who made them, and when they were made.
- **Pull Requests:** Developers can create pull requests on GitHub, which allow them to suggest changes to a project and get feedback from other developers before those changes are merged into the main codebase.
- **Issue Tracking:** GitHub provides a robust issue tracking system, which allows developers to track bugs, feature requests, and other issues related to their projects.

- **Collaborative Tools:** GitHub provides a number of tools for collaboration, such as wikis, project boards, and team management features. These tools make it easy for developers to work together on projects and keep everyone on the same page.
- **Community Features:** GitHub is a large online community of developers, and the platform provides a number of community features such as social coding, notifications, and forums, which allow developers to interact with each other and learn from one another.
- **Security:** GitHub takes security seriously and provides a number of features to help protect code repositories and user accounts, such as two-factor authentication, encryption, and vulnerability scanning.
- **Overall, GitHub is a powerful platform that provides a wide range of tools and features to help developers collaborate on software projects and build better software together.**

The screenshot shows the GitHub Actions interface for the repository `VINAYAKBISWAGAR / SolarEnergyPrediction`. The sidebar on the left lists various actions, with 'All workflows' selected. The main content area shows a list of workflow runs for the 'pages build and deployment' workflow. The runs are listed in a table with columns for Event, Status, Branch, and Actor.

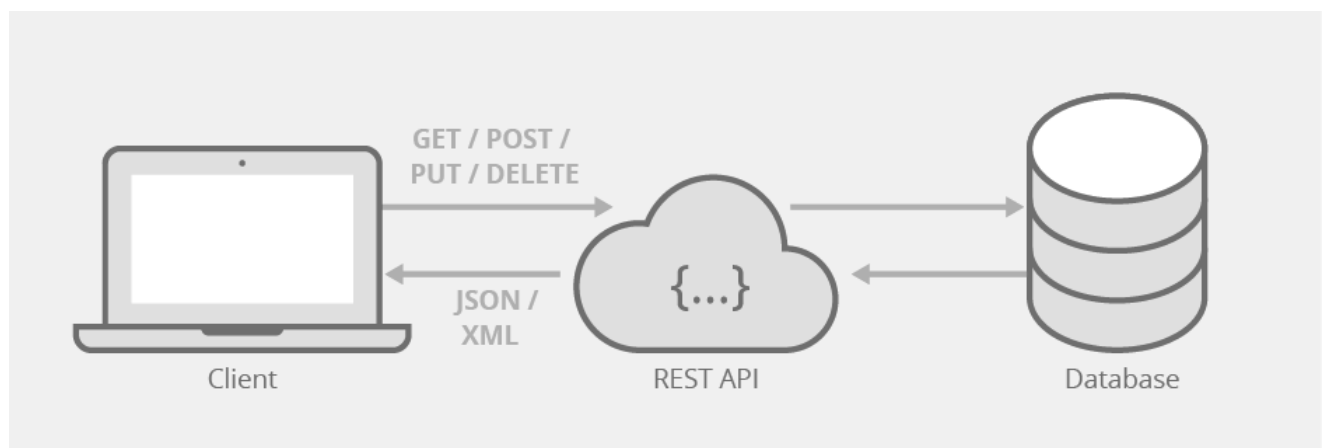
Event	Status	Branch	Actor
pages build and deployment	Success	pages-build-deployment #3: by VINAYAKBISWAGAR	2 weeks ago 58s
pages build and deployment	Success	pages-build-deployment #2: by VINAYAKBISWAGAR	2 weeks ago 1m 5s
pages build and deployment	Success	pages-build-deployment #1: by VINAYAKBISWAGAR	2 weeks ago 55s

Flask

Flask is a popular web development framework for Python that is used to build web applications. Flask is a lightweight and flexible framework that is easy to use and highly customizable. Flask is based on the Werkzeug WSGI toolkit and the Jinja2 template engine, which makes it highly extensible and versatile.

Some of the key features of Flask include:

- **Routing:** Flask provides a powerful routing system that enables developers to map URLs to specific views or functions.
- **Templating:** Flask comes with the Jinja2 templating engine, which allows developers to create HTML templates that can be dynamically rendered with data from the application.
- **Debugging:** Flask provides a built-in debugger that makes it easy to identify and fix errors in the application code.
- **Extensions:** Flask has a large number of extensions available that can be used to add additional functionality to the framework, such as support for databases, authentication, and caching.
- **Integration:** Flask can be easily integrated with other Python libraries and frameworks, such as NumPy, Pandas, and Tensor Flow.
- **Deployment:** Flask applications can be easily deployed on a variety of platforms, including cloud hosting services like Amazon Web Services and Microsoft Azure.
- **Security:** Flask provides built-in support for secure coding practices, such as protecting against cross-site scripting (XSS) and SQL injection attacks.



<https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FFlask>

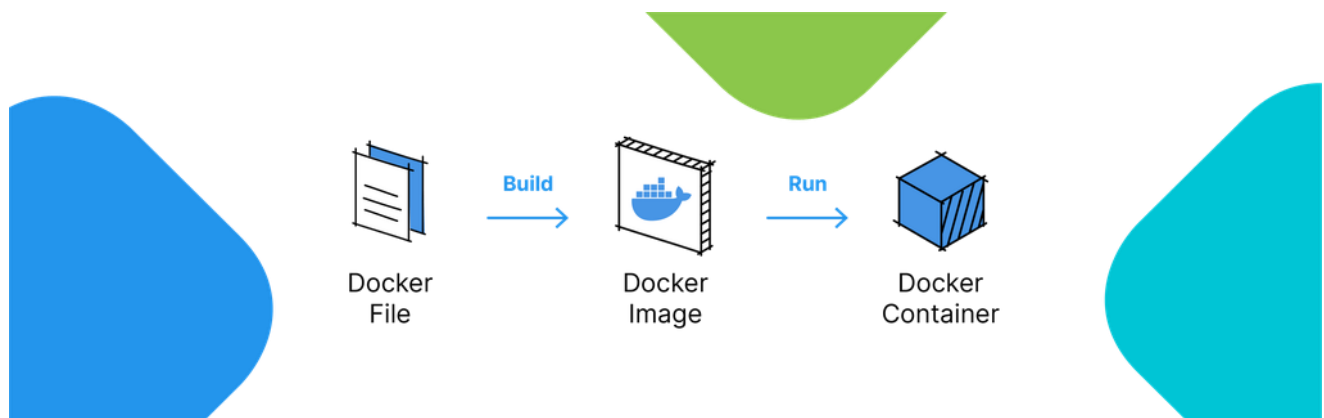
Overall, Flask is a flexible and powerful web development framework that can be used to build a wide range of web applications, from small personal projects to large-scale enterprise applications. Its simplicity, flexibility, and extensibility make it a popular choice among developers.

Docker

Docker is a containerization platform that allows developers to package their applications and dependencies into a portable container. Containers are lightweight and consume fewer resources than traditional virtual machines, which makes them ideal for running multiple applications on a single server. Docker is based on the idea of containerization, which is a way of packaging an application and all its dependencies into a single container that can be run on any platform that supports Docker.

Some of the key features of Docker include:

- **Portability:** Docker containers are highly portable, which means that they can be easily moved between different environments without requiring any changes to the application code.
- **Scalability:** Docker allows applications to be easily scaled up or down, based on the demand for resources.
- **Consistency:** Docker ensures that the application runs consistently across different environments, by packaging all the necessary dependencies into the container.
- **Security:** Docker provides built-in security features, such as container isolation and image signing that help to protect the application from attacks.
- **Efficiency:** Docker containers are lightweight and consume fewer resources than traditional virtual machines, which makes them ideal for running multiple applications on a single server.
- **Collaboration:** Docker makes it easy for developers to collaborate on projects, by providing a common platform for packaging and sharing applications and dependencies.
- **Integration:** Docker can be easily integrated with other tools and platforms, such as Kubernetes and Amazon Web Services, to provide a complete container-based solution for application deployment and management.



<https://www.google.com/url?sa=i&url=https%3A%2F%2Fbikramat.medium.com%2Fdoc>

Overall, Docker is a powerful platform for containerization that provides many benefits for developers, including portability, scalability, consistency, security, efficiency, collaboration, and integration. Its popularity has led to a large ecosystem of tools and services that make it easy to use Docker in a wide range of applications and environments.

Amazon Web Services

Amazon Web Services (AWS) is a comprehensive cloud computing platform that offers a wide range of services for businesses and individuals.

Some of the key features of AWS include:

- **Scalability:** AWS allows users to easily scale their applications up or down based on the demand for resources. This is achieved by running multiple instances of the same application, which enables it to handle more requests and users.
- **Reliability:** AWS is known for its high reliability, thanks to its multiple data centers located around the world. This provides high availability and redundancy for applications and data. AWS also provides backup and disaster recovery services to protect against data loss and system failures.
- **Security:** AWS has built-in security features, including network isolation, access control, and encryption, to protect against cyber-attacks and data breaches. AWS complies with several industry standards and regulations, such as SOC 2, PCI DSS, and HIPAA, making it a trusted platform for sensitive data.
- **Cost-effectiveness:** AWS offers a pay-as-you-go pricing model, allowing users to only pay for the resources they use. This makes it a cost-effective solution for businesses of all sizes.
- **Flexibility:** AWS provides a wide range of services, including compute, storage, database, analytics, machine learning, and more. This allows businesses to choose the services that best meet their needs.
- **Innovation:** AWS is constantly introducing new services and features to stay ahead of the competition and meet the evolving needs of its customers.



[https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%](https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FAmazon+Web+Services)

Overall, AWS is a powerful and versatile platform that offers a range of benefits to businesses and individuals alike. Its scalability, reliability, security, cost-effectiveness, flexibility, and innovation make it a top choice for cloud computing.

Conclusion

Some useful findings, which are observed in other studies, are for instance substantial improvements for long-term forecasting when using numerical weather predictions (NWP) as input data. Furthermore, modeling future cloud positions with satellite-based data have improved short-term solar PV energy output forecasting. Finally, model accuracy tends to vary depending on the climatic condition of the forecasting location. As of this, a model is likely to perform better in one region than when trained on multiple sites simultaneously. Similarly, as climatic conditions can vary over a yearly cycle, a model may perform better when trained on one weather season rather than several

Reference:

- 1 Aneela Zameer, Farah Shahid², Mudasser Afzal, Muhammad Hassan, “Intelligent forecast models for daily solar energy prediction”, Research gate- October 2020**
- 2 K. Anuradha¹, Deekshitha Erlapally , G. Karuna , V. Srilakshmi , K. Adilakshmi Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques, E3S Web of Conferences 309, 01163 ICMED 2021**
- 3 Bouchaib Zazoum Solar photovoltaic power prediction using different machine learning methods, 2021 8th International Conference on Power and Energy Systems Engineering (CPESE 2021)**
- 4 C. Vennila, Anita Titus, T. Sri Sudha, U. Sreenivasulu, N. Pandu Ranga Reddy, K. Jamal, Dayadi Lakshmaiah, P. Jagadeesh and Assefa Belay Forecasting Solar Energy Production Using Machine Learning, Hindawi International Journal of Photoenergy Volume 2022, Article ID 7797488, 7 pages <https://doi.org/10.1155/2022/7797488>**