Weather Prediction

By: Nidhi Pravinbhai Patel

Predicting Daily Weather Summary Using Machine Learning

Step- 1 : Prototype Selection

Objective: Develop a machine learning model to predict the daily weather summary based on historical weather data. The model should accurately classify the daily weather summary using various features such as 'Formatted Date', 'Summary', 'Precip Type', 'Temperature (C)', 'Apparent Temperature (C)', 'Humidity', 'Wind Speed (km/h)', 'Wind Bearing (degrees)', 'Visibility (km)', 'Loud Cover','Pressure (millibars)', 'Daily Summary'.

Market/ Customer / Business need assessment:

Consumer Needs:

1. **Personal Safety and Convenience**:
   - **Daily Planning**: Accurate weather forecasts help with commuting, outdoor activities, and travel plans.
   - **Health and Safety**: Alerts for extreme weather conditions to take necessary precautions.
   - **Customization**: Personalized forecasts based on location and activities.
2. **Technological Integration**:
   - **Smart Devices and IoT**: Integration with home devices (e.g., thermostats, sprinklers) for automated responses.
   - **Mobile Applications**: High demand for user-friendly, real-time weather apps.

Business Needs:

1. **Industry-Specific Applications**:
   - **Agriculture**: Decisions on planting, irrigation, and harvesting.
   - **Transportation and Logistics**: Safety, route optimization, and delay reduction.
   - **Construction**: Planning and scheduling to avoid weather-related delays.
   - **Energy Sector**: Forecasting energy production and managing supply.
2. **Risk Management and Mitigation**:
   - **Insurance**: Risk assessment and claim management.
   - **Disaster Management**: Preparation and response to natural disasters.
3. **Economic Efficiency**:
   - **Operational Cost Savings**: Optimizing operations and reducing costs.
   - **Resource Allocation**: Efficient scheduling and productivity enhancement.

Demand Analysis

- **Consumer Demand**: High engagement with weather apps, expecting accurate forecasts and additional features.
- **Business Demand**: Industries increasingly adopt advanced weather prediction models for better decision-making and operational efficiency.

Target Specification and Characterization (customer characteristic):

1. **Consumers**
   - Needs: Accurate daily forecasts, severe weather alerts, easy-to-use apps.
   - Preferences: Personalized updates, smart device integration, visual data.
2. **Businesses**
   - Industries: Agriculture, transportation, construction, energy, insurance.
   - Needs: Accurate forecasts for planning, real-time data, customizable reports.
   - Preferences: API access, detailed analytics, reliable performance.
3. **Government/Public Sector**
   - Agencies: Meteorological, disaster management, environmental monitoring.
   - Needs: Precision for safety, large-scale analysis, accurate reports.
   - Preferences: Advanced tools, real-time monitoring, data sharing.
4. **Technological Integrators**
   - Companies: Smart home, IoT, app developers.
   - Needs: Easy API integration, real-time streaming, scalability.
   - Preferences: Detailed documentation, flexible data formats, high availability.

Various models have different performance according to its features , algorithm used. Following analysis is done based upon the above information.

IBM Watson: Best for advanced analytics and large-scale business applications but at a higher cost.

AccuWeather: High accuracy and excellent for consumer use with comprehensive apps but less customizable for businesses.

Dark Sky: Superior for hyper-local forecasts and user experience, ideal for individual consumers, less suited for businesses.

Weather Underground: Great for community-based reporting and localized data, variable data quality.

NOAA: Reliable, authoritative, and free, but not as user-friendly or customizable.

OpenWeatherMap: Cost-effective and scalable with extensive API support, variable accuracy.

External Search:

- - - - - - - - - - - - - x

The dataset I used in the model making program is a small dataset that I created randomly using the features Apparent Temperature (C),Precip type,temperature,wind speed ,Humidity,Wind Bearing (degrees) , Visibility (km).

The activities of many primary sectors depend on the weather for production, e.g. farming. The climate is changing at a drastic rate nowadays, which makes the old weather prediction methods less effective and more hectic. To overcome these difficulties, improved and reliable weather prediction methods are required. These predictions affect a nation's economy and the lives of people. To develop a weather forecasting system that can be used in remote areas is the main motivation of this work. The data analytics and machine learning algorithms, such as random forest classification, are used to predict weather conditions. In this paper, a low-cost and portable solution for weather prediction is devised.

https://janaksenevirathne.medium.com/building-a-weather-prediction-model-with-machine-learning-a-step-by-step-guide-

Journals: https://ieeexplore.ieee.org/abstract/document/8938211

https://ieeexplore.ieee.org/abstract/document/8441679

Benchmarking Alternate Products:

Comparing Commercial Weather Services: Assess the accuracy and reliability of forecasts from different commercial providers (e.g., IBM Weather Company vs. AccuWeather) for predicting severe weather events.

Evaluation of Open-Source Solutions: Benchmark open-source frameworks like MetPy or NOAA NCEP models against proprietary solutions to determine their suitability for research or educational purposes.

Integration with Operational Systems: Evaluate the ease of integration and operational deployment of each product within existing infrastructure, such as aviation weather services or agricultural advisory systems.

Applicable Patents:

1. **US Patent 10,332,169** - Uses machine learning for weather forecasting based on historical data analysis.
2. **US Patent 9,837,095** - Applies deep learning to predict weather patterns from real-time and historical data.
3. **US Patent 9,631,048** - Predicts weather conditions using machine learning models trained on diverse atmospheric data sources.
4. ClimateAi Has U.S. Patent Granted for GenAI-Based Approach Applied to Weather Forecasting

Products:

1. **IBM Weather Company**: Offers machine learning-driven weather prediction services for businesses and consumers.
2. **AccuWeather**: Provides real-time weather updates and forecasts using machine learning algorithms.
3. **The Weather Company (IBM)**: Delivers personalized weather forecasts through apps and APIs, leveraging machine learning for accuracy.

Applicable Regulations (government and environmental regulations imposed by countries):

1. **Data Protection and Privacy Laws**:
    - **GDPR, CCPA**: Ensure personal data protection and user control over data used in weather prediction models.
2. **Meteorological Data Regulations**:
    - Govern how meteorological data can be collected, shared, and used, often by national meteorological services.
3. **Ethical AI/ML Use**:
    - Guidelines emphasize fairness, transparency, and accountability in using AI/ML for weather prediction.
4. **Environmental Impact**:
    - Consider environmental regulations for data collection and modeling activities, especially related to climate change and air quality.

By adhering to these regulations, ML models for weather prediction can operate legally and ethically, contributing effectively to environmental and public safety efforts.

Applicable constraints(need for space,expertise,budet) for weather prediction machine learning model:

1. **Data Requirements**:
    - **Space**: Need for extensive storage for large weather datasets.
    - **Expertise**: Skilled data scientists and meteorologists.
    - **Budget**: High costs for data acquisition and computing resources.
2. **Computational Resources**:
    - **Space**: Requires significant computing power for data processing.
    - **Expertise**: Knowledge in cloud computing or high-performance computing.
    - **Budget**: Expenses for cloud services or hardware.
3. **Model Complexity and Performance**:
    - **Space**: Complex algorithms and model tuning for accuracy.
    - **Expertise**: Expert knowledge in machine learning and model optimization.
    - **Budget**: Investment in model development and testing.
4. **Regulatory and Ethical Considerations**:
    - **Space**: Compliance with data privacy laws (e.g., GDPR, CCPA).
    - **Expertise**: Legal and ethical expertise for responsible AI deployment.
    - **Budget**: Costs for regulatory compliance and ethical frameworks.
5. **Operational Constraints**:
    - **Space**: Real-time processing capabilities for timely forecasts.
    - **Expertise**: Operational integration of ML models into weather systems.
    - **Budget**: Maintenance and support costs.

Managing these constraints ensures effective deployment and operation of ML models for weather prediction while meeting technical, regulatory, and operational requirements.

a. Feasibility:

- Short-Term Development (2-3 Years)
- **Data Accessibility**: Weather data is readily available and commonly used for such models.
- **Established Tools**: Technologies and libraries (e.g., Python, scikit-learn) are well-documented and widely used.
- **Prototype Speed**: Quick development of a working model and iterative improvements are feasible within this timeframe.

b. Viability:

- Long-Term Relevance (20-30 Years)
- **Ongoing Demand**: Weather forecasting is essential across multiple sectors and will remain relevant due to climate changes.
- **Adaptability**: The model can evolve with advancements in technology and new data sources, maintaining its value over time.
- **Scalability**: The model can be scaled and updated to handle more data and provide more detailed predictions.

c. Monetization:

- Direct Revenue Generation
- **Subscription Models**: Offer premium services or advanced features through subscriptions.
- **API Access**: Charge for API usage, enabling integration into other applications.
- **Custom Solutions**: Provide tailored services to industries that need specialized weather forecasts.

This approach ensures that the weather predictive model is feasible to develop, viable long-term, and directly monetizable.

Step - 2: Prototype Development

1. **Data exploration:** The code starts by looking at the data, checking what it contains, and summarizing its features like Formatted Date, Summary, Precip Type, Temperature (C),Apparent Temperature (C)'Humidity,Wind Speed (km/h),Wind Bearing (degrees), Visibility (km), Loud Cover,Pressure (millibars) and based on these parameters it predicts the value for Daily Summary.

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)	Daily Summary
0	2006-04-01 00:00:00.000 +0200	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251.0	15.8263	0.0	1015.13	Partly cloudy throughout the day.
1	2006-04-01 01:00:00.000 +0200	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259.0	15.8263	0.0	1015.63	Partly cloudy throughout the day.
2	2006-04-01 02:00:00.000 +0200	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204.0	14.9569	0.0	1015.94	Partly cloudy throughout the day.
3	2006-04-01 03:00:00.000 +0200	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269.0	15.8263	0.0	1016.41	Partly cloudy throughout the day.
4	2006-04-01 04:00:00.000 +0200	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259.0	15.8263	0.0	1016.51	Partly cloudy throughout the day.
...	...	...	...	...	...	...	...	...	...	...	...	...
96448	2016-09-09 19:00:00.000 +0200	Partly Cloudy	rain	26.016667	26.016667	0.43	10.9963	31.0	16.1000	0.0	1014.36	Partly cloudy starting in the morning.

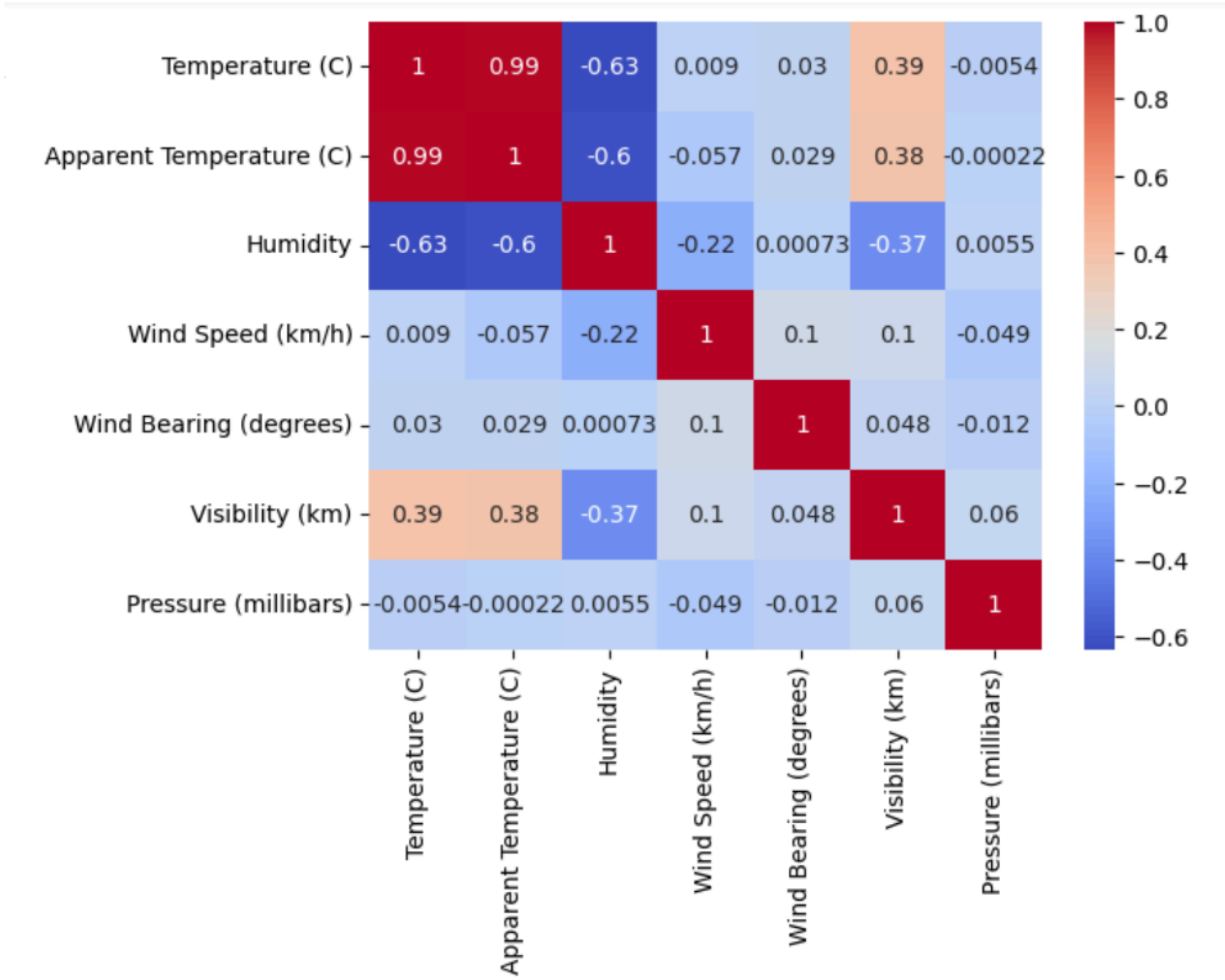
2. **Visualize the relationship:** It creates a plot that shows the relationship between different features.

```
data.shape
```

```
(96453, 12)
```

```
sns.heatmap(data[['Temperature (C)', 'Apparent Temperature (C)',
'Humidity', 'Wind Speed (km/h)', 'Wind Bearing (degrees)',
'Visibility (km)', 'Pressure (millibars)']].corr(),

            annot=True,cmap='coolwarm')
```



3. Preparing Data: After understanding the data, it prepares for analysis by selecting the important features and scaling them as categorical and numerical features.

```
# Identify categorical and numerical features

categorical_features = ['Summary', 'Precip Type']

numerical_features = ['Temperature (C)', 'Apparent Temperature (C)',
'Humidity', 'Wind Speed (km/h)', 'Wind Bearing (degrees)',
'Visibility (km)', 'Pressure (millibars)']



# Drop 'Formatted Date' and 'Loud Cover' since they are not useful
for prediction

data = data.drop(['Formatted Date', 'Loud Cover'], axis=1)



# Handle missing values if any

data = data.dropna()



# Encode categorical features

data_encoded = pd.get_dummies(data, columns=categorical_features,
drop_first=True)



# Label encode the target variable 'Daily Summary'

label_encoder = LabelEncoder()

data_encoded['Daily Summary'] =
label_encoder.fit_transform(data_encoded['Daily Summary'])



# Split data into features and target
```

```
X = data_encoded.drop('Daily Summary', axis=1)

y = data_encoded['Daily Summary']



print(X.shape,y.shape)
```

4. Training a model: Random Forest Classifier is favored in weather prediction for its ability to handle complex interactions among weather variables, mitigate overfitting with ensemble learning, and provide insights into feature importance critical for accurate forecasting.

```
# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Normalize numerical features
scaler = StandardScaler()
X_train[numerical_features] = scaler.fit_transform(X_train[numerical_features])
X_test[numerical_features] = scaler.transform(X_test[numerical_features])

# Initialize and train the model
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Evaluate the model on the test set
y_pred = model.predict(X_test)
```

Concept of development:

- - - - - - - - - - - - - X

Brief summary of product or service

A weather predicting product/service will utilize machine learning to provide accurate and timely forecasts for various weather parameters such as temperature, precipitation, and wind conditions. It will integrate historical and real-time data sources, apply advanced algorithms for prediction, and offer customizable features tailored to specific user needs. The service aims to enhance decision-making in sectors like agriculture, transportation, and emergency management, ensuring reliable insights and operational efficiency in response to weather fluctuations.

Data Integration: Collects data from meteorological stations, satellites, and IoT devices for historical and real-time weather variables.

Machine Learning Algorithms: Uses advanced algorithms like Random Forests or LSTM to predict weather conditions accurately.

Customizable Features: Offers personalized alerts, interactive dashboards, and APIs for integration into existing systems.

Sector-Specific Applications: Targets agriculture, transportation, energy, and emergency management for decision support.

Accuracy and Reliability: Focuses on high accuracy through continuous model refinement and validation against ground truth data.

User Interface: Provides user-friendly interfaces via web, mobile apps, and APIs for easy access to weather insights.

Support and Maintenance: Includes regular updates, proactive monitoring, and responsive customer service.

Step - 3 : Business Model (Monetization Idea)

- **Subscription-Based Services**: Offer plans for weather forecasts and historical data, with tiered pricing based on forecast frequency and detail.
- **APIs and Data Licensing**: Provide APIs for integrating weather data into applications, charging based on usage metrics like data volume or API calls.
- **Custom Solutions and Consulting**: Offer tailored consulting services and develop custom weather models for industries like agriculture, logistics, and construction.

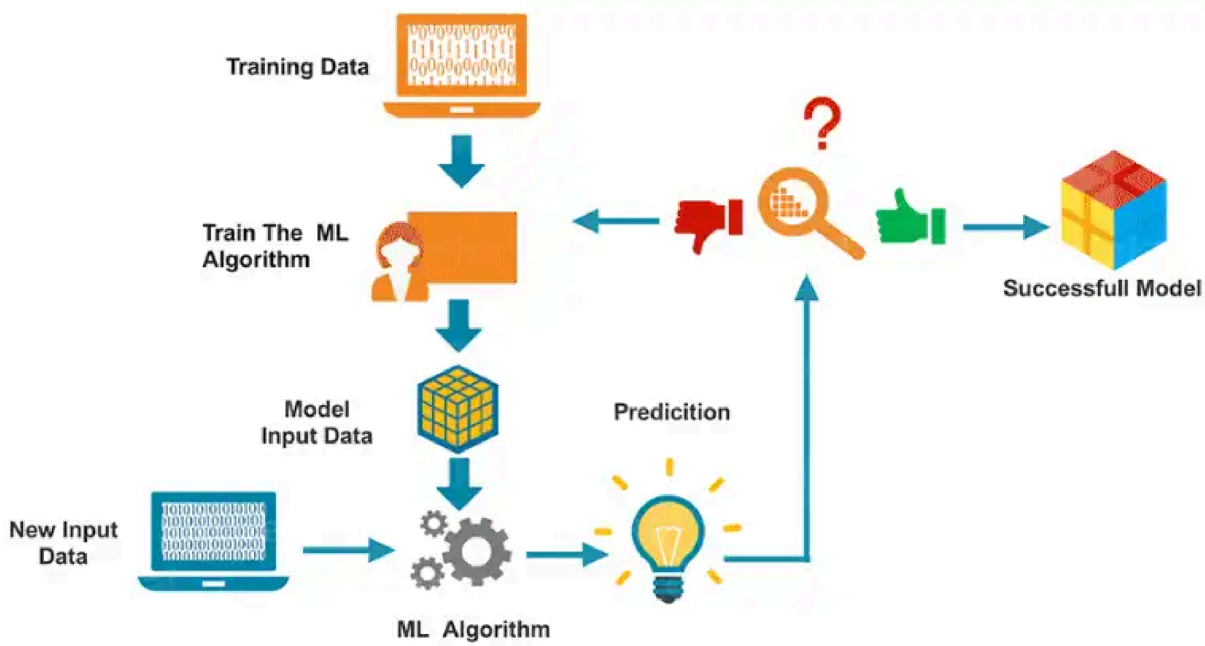
- **Advertising and Sponsorships**: Partner with advertisers for targeted weather-related ads and explore sponsorship opportunities in media or weather events.
- **Premium Features and Data Insights**: Charge for access to advanced analytics, historical data analysis, and specialized weather reports.

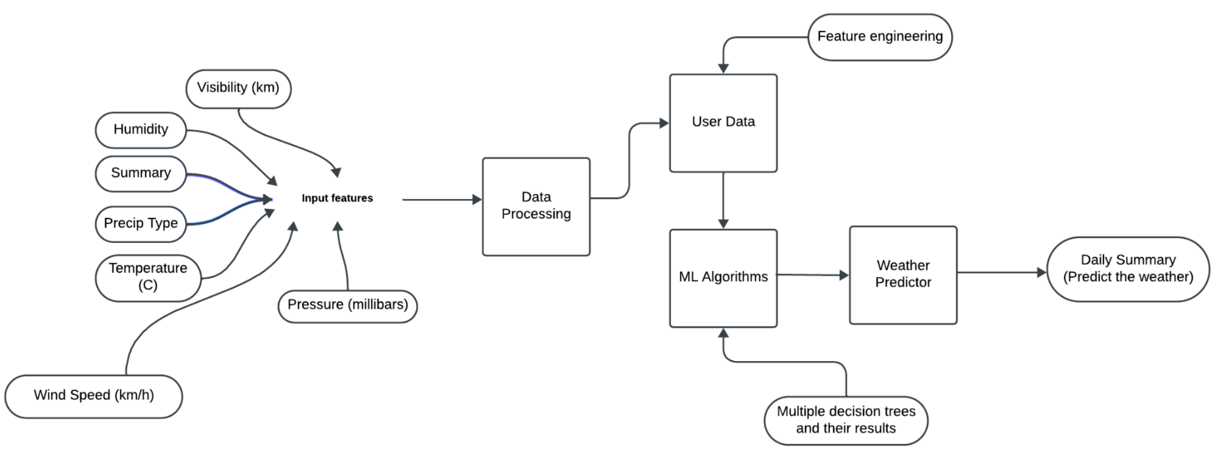
Identify target markets and customize offerings to meet their specific needs,Emphasize accuracy, reliability, and actionable insights to differentiate from free or less precise alternatives,Consider expenses for data acquisition, infrastructure (e.g., cloud computing), R&D, and regulatory compliance,Utilize online platforms, industry conferences, partnerships, and direct sales efforts to reach potential customers.

By leveraging these monetization strategies, a weather prediction model can generate sustainable revenue while providing valuable services and insights to a wide range of industries and stakeholders.

Abstract of final product prototype:

Weather prediction is crucial for various applications, from agriculture and transportation to disaster management and energy optimization. This study explores the use of Random Forest, an ensemble learning technique, for accurate weather forecasting. Historical weather data comprising variables such as temperature, humidity, wind speed, and precipitation are utilized. The Random Forest model is trained on this data to predict future weather conditions, demonstrating robustness in handling complex relationships and providing insights into feature importance. Evaluation metrics such as accuracy and mean squared error validate the model's effectiveness in predicting weather patterns, highlighting its potential for real-time applications and decision support systems in diverse industries.





1. Data Acquisition

- **Collect Data**: Obtain weather data from a CSV file with columns like 'Summary', 'Precip Type', 'Temperature (C)', etc.

2. Data Preprocessing and Model Training

- **Handle Missing Data**: Address missing values by dropping or filling them.
- **Feature Selection**: Choose relevant features (e.g., temperature, humidity).

- **Encode Categorical Variables**: Convert categorical data (e.g., 'Summary', 'Precip Type') to numerical format using techniques like one-hot encoding.
- **Split Data**: Divide into training and testing sets.
- **Normalize/Standardize**: Scale numerical features for consistency.

3. Model Building

- **Select Model**: Use a RandomForestClassifier for the classification task.
- **Train Model**: Fit the model on the training data to learn patterns.

4. Application Development

- **Develop UI**: Create a web app with frameworks like Streamlit or Flask for user input.
- **Handle Input**: Collect user inputs and preprocess them (e.g., encoding, scaling).

5. Predicting Output

- **Make Predictions**: Use the trained model to predict the weather summary based on user inputs.
- **Display Results**: Show the predicted summary to the user through the web app interface.

Product Details :

\- - - - - - - - - - - x

How Does It Work?

I have created a Web App for this project using a small dataset of weather and giving the result to the user with a value of what the weather can be. The app calculates the user's input values for the features and uses the RandomForestClassifier model to predict the future weather by random forest.

Step - 4 : Financial Modeling

A. Identify which Market your product/service will be launched into

1. Weather Services
2. Agriculture and Farming
3. Travel and Transportation
4. Emergency Management and Disaster Response
5. Energy Sector
6. Retail and Consumer Goods

B. Collect some data /statistics regarding that Market Online.

1. Weather Services Market

- Description: Providers of weather forecasts and information.
- Target Customers: Meteorological agencies, weather apps, media outlets.
- Opportunity: Enhanced accuracy and specialized forecasting services.

2. Agriculture and Farming

- Description: Uses weather data for crop management and risk mitigation.
- Target Customers: Farmers, agricultural businesses, agritech companies.
- Opportunity: Tailored forecasts for crop management and irrigation.

3. Travel and Transportation

- Description: Weather impacts travel logistics and safety.
- Target Customers: Airlines, shipping companies, logistics firms, travel agencies.
- Opportunity: Real-time forecasts to improve scheduling and safety.

4. Emergency Management and Disaster Response

- Description: Relies on weather information for disaster preparedness.
- Target Customers: Government agencies, emergency services, disaster response organizations.
- Opportunity: Detailed forecasts and alerts for disaster management.

5. Energy Sector

- Description: Weather data optimizes energy operations, especially renewable sources.
- Target Customers: Solar and wind energy providers, utility companies.
- Opportunity: Predictions for energy production and grid management.

6. Retail and Consumer Goods

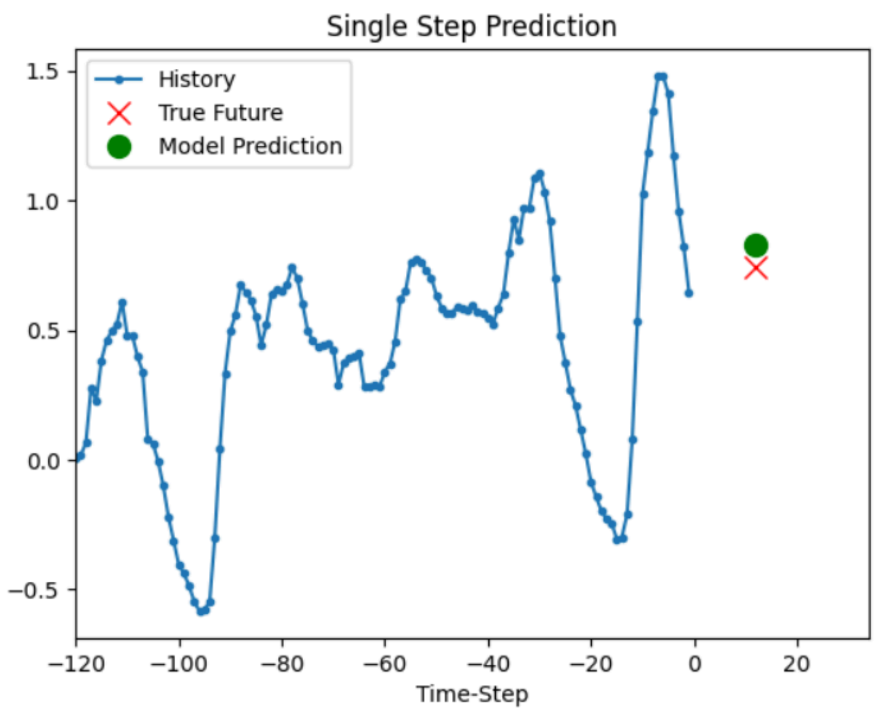
- Description: Uses weather data for inventory and supply chain management.
- Target Customers: Retailers, manufacturers, e-commerce businesses.
- Opportunity: Forecasts for inventory planning and demand forecasting.

C. Perform forecasts/predictions on that Market using regression models or time series forecasting (alternately collect existing Statistics if you are unable to find appropriate data or perform time series)

Time Series forecasting for weather prediction using Jena Climate dataset recorded by the Max Planck Institute for Biogeochemistry. The dataset consists of 14 features such as temperature, pressure, humidity etc, recorded once per 10 minutes.

Location: Weather Station, Max Planck Institute for Biogeochemistry in Jena, Germany

Time-frame Considered: Jan 10, 2009 - December 31, 2016



D. Financial Equation

Exponential Smoothing models are used for forecasting by giving exponentially decreasing weights to past observations. The simplest form is:

$$y_t = \alpha x_t + (1 - \alpha) y_{t-1}$$

Where:

- α is the smoothing parameter (0 < α < 1).
- xt is the observed value at time t.
- yt−1 is the forecasted value from the previous period.

For more sophisticated models, consider the Holt-Winters method which includes trend and seasonality components:

$$y_t = (l_{t-1} + b_{t-1}) + s_{t-L} + \epsilon_t$$

- lt: Level component.
- bt: Trend component.
- st−L: Seasonal component.

Code Implementation/Validation on Small Scale:

\- x

This is a github link for the project implementation

https://github.com/nidhi-158/second-project/blob/main/Weather_Prediction.ipynb

Conclusion:

\- - - - - - - x

In conclusion, the fusion of machine learning and meteorology holds the promise of transforming reliable ability to anticipate and respond to weather phenomena. As we continue to refine and innovate in this space, the potential for improving the accuracy and timeliness of weather forecasts becomes limitless. Our journey into predicting tomorrow's weather unveils not only the capabilities of machine learning but also its profound impact on shaping a more resilient and informed society.

These methods are extremely easy to adopt as they don't require any specific computational power like Deep Learning methods (RNN, CNN …).It is important to consider that we only have examined monthly average values while it may be interesting to consider daily values too and have daily predictions.

References:

- - - - - - - - x

https://towardsdatascience.com/weather-forecasting-with-machine-learning-using-python-55e90c346647

https://janakseneviratne.medium.com/building-a-weather-prediction-model-with-machine-learning-a-step-by-step-guide-9eaf768171be

https://keras.io/examples/timeseries/timeseries_weather_forecasting/