Article

Research and Application of Intelligent Weather Push Model Based on Travel Forecast and 5G Message

Yuan Yuan [1], Fengchen Fu [2,*], Yaling Li [1], Yao Xing [3], Lei Wang [2], Hao Zheng [1] and Wei Ye [1]

[1] Sichuan Meteorological Service Centre, Chengdu 610072, China; yuanyuan384442@cma.cn (Y.Y.)
[2] Shaanxi Meteorological Information Centre, Xi'an 710014, China
[3] Huyi District Meteorological Bureau, Xi'an 710399, China
[*] Correspondence: fuliya005@foxmail.com

Abstract: In the realm of daily activity planning, precise weather forecasting services hold paramount significance. However, the prevalent dissemination of weather forecasts through conventional channels like radio, television, and the internet often yields only generalized regional predictions. This limitation contributes to diminished forecast reach, inadequate accuracy, and a lack of individualization, thwarting the effective distribution of meteorological insights and inhibiting the fulfillment of personalized forecast demands. Addressing these concerns, our study proposes a personalized weather forecasting approach that harnesses machine learning techniques and leverages the 5G messaging platform. By amalgamating projected user travel data, we augment personalized weather reports and extend user coverage to achieve tailored, timely, and high-quality weather services. Concretely, our research commences with an extensive analysis of large-scale user travel behavior data to extract pertinent travel attributes. Subsequently, we construct a user's future location prediction model—dubbed the Loc-PredModel—by employing the Extreme Gradient Boosting (XGBoost) algorithm to forecast users' trip destinations and arrival times. Anchored in the anticipated outcomes of user travel behavior, personalized weather data reports are formulated. Experimental results underscore the Loc-PredModel's remarkable predictive prowess, demonstrating a root mean squared error (RMSE) value of 0.208 and a coefficient of determination (R^2) value of 0.935, affirming its efficacy in prognosticating users' trip destinations and arrival times. Furthermore, our 5G message-driven platform, rooted in intelligent personalized meteorological services, underwent testing within Chengdu city and garnered positive user feedback. Our research effectively surmounts the limitations of conventional weather forecasting platforms by furnishing users with more precise and customized weather information predicated on behavioral analysis and the 5G information ecosystem. This study not only advances the theoretical groundwork of intelligent meteorology but also offers invaluable insights and guidance for future advancement. By providing users with a more personalized and timely intelligent meteorological service experience, our approach exhibits transferability, with the research methodology and model potentially extendable nationwide or even on a larger scale beyond the study's Chengdu-based scope.

Keywords: behavior prediction; personalized delivery; intelligent meteorological services; 5G messaging platform; machine learning

1. Introduction

Weather has a significant impact on people's daily lives and work [1]. Accurate meteorological information can assist individuals in better organizing their work and life, thereby enhancing productivity [2] and improving the quality of life. As the pace of life accelerates, there is an increasing demand for the accuracy of weather forecasts [3], with users' personalized requirements becoming more prominent. Despite the breakthroughs achieved in intelligent and automated meteorological monitoring technology [4], the quality and quantity of meteorological data have improved significantly, and the accuracy and

resolution of forecasts continue to improve [5]. However, they also face the challenge of information overload brought about by the widespread proliferation of the internet [6]. Particularly in terms of the dissemination channels for meteorological data, traditional media such as broadcast television and the internet still dominate [7]. However, these channels often only provide rough forecasts over a large regional scope, failing to meet modern society's demand for personalized, precise, and timely meteorological information. For instance, different locations within the same administrative region of a city can experience vastly different weather conditions. Higher geographical resolution services can offer more accurate meteorological insights. This situation results in low forecast reach, inadequate precision, and a lack of personalized services, thereby restricting the practical application of meteorological data in everyday life.

Therefore, personalized weather recommendations have emerged as a novel research direction. Leveraging techniques like machine learning, these studies utilize historical data to offer more accurate and personalized weather recommendations. While such research contributes to enhancing prediction accuracy and user experience, the limitations of personalized weather recommendations persist, unable to address the issue of timely information delivery in meteorology.

To address the aforementioned issues, researchers have been striving to enhance the accuracy and personalization of these services [8–10]. In this pursuit, there has been accelerated technological innovation to ensure precise monitoring, forecasting, and detailed services [11]. Government reports also emphasize the strengthening of disaster prevention and emergency capabilities through improved meteorological services. One key aspect involves advancing core technologies such as refining weather mechanisms and enhancing numerical forecasting models [12]. Through the combination of these efforts, researchers are committed to improving the quality, timeliness, and personalization of meteorological services, thereby contributing to the overall well-being and safety of society.

However, despite the significant progress made in the field of meteorological services, current research primarily focuses on enhancing meteorological services but lacks an in-depth exploration from the perspective of user behavior analysis to improve the accuracy of meteorological services. Personalized weather forecasting research that is based on user behavior analysis and takes a user-centric approach is currently relatively scarce. This research perspective holds great significance. Leveraging technologies such as machine learning, personalized weather forecasting based on user behavior analysis can tailor weather forecasts to individual users by analyzing their historical behaviors, preferences, and backgrounds, better meeting their unique needs and patterns. This forward-looking research direction holds the potential to open up new avenues for enhancing the accuracy and personalization level of meteorological services. Delving into the correlation between user behavior and meteorology will provide a fresh perspective for personalized weather forecasting, thereby further elevating the precision and personalization level of meteorological services.

Therefore, this study aims to comprehensively consider the strengths and weaknesses of these models and integrate them with the 5G messaging platform to establish a refined and efficient personalized meteorological service system. By deeply analyzing user travel behavior, integrating multi-model predictions, and leveraging the rich media and interactivity features of the 5G messaging platform, personalized, timely, and high-quality weather forecasts can be achieved to meet diverse user demands. The selection of the 5G messaging platform is driven by its superior bandwidth, lower latency, and enhanced capacity compared to 4G networks, enabling the delivery of rich media content and facilitating more interactive user experiences. While the 4G platform can also support similar services, the 5G platform offers distinct advantages in terms of delivering immersive and engaging meteorological information, thereby significantly enhancing the quality and efficiency of meteorological information services.

This research holds significant theoretical and practical value. Theoretically, it explores novel personalized weather recommendation methods by leveraging the rich media and

interactivity features of the 5G messaging platform, addressing the limitations of traditional approaches. Practically, it offers new communication channels and interaction methods to improve the reach and user satisfaction of meteorological information, promoting intelligent, personalized, and accurate meteorological information services.

2. Related Work

In the domains of personalized weather forecasting and smart meteorological services, previous research has explored various approaches, yet limitations persist. Many methods overly rely on traditional meteorological data and models, struggling to meet personalized demands. Some studies overlook user behaviors and preferences, focusing excessively on data analysis and prediction. This study aims to bridge these research gaps by analyzing user travel behavior data to extract travel patterns and preferences. We employ a future location prediction model based on the Extreme Gradient Boosting (XGBoost) algorithm to accurately predict users' upcoming destinations. Combining individual user characteristics, personalized meteorological data reports are generated. Leveraging the 5G messaging platform, we achieve a tailored smart meteorological service that caters to diverse users.

Compared to prior research, our approach offers innovations: a focus on user behavior to better address practical needs, the application of machine learning for more accurate predictions, and the utilization of the 5G messaging platform for timely delivery, enhancing both timeliness and personalization.

Integrating artificial intelligence, big data, and cloud computing, intelligent weather push technology offers real-time and precise weather information based on user requirements and contexts. Key steps encompass data collection, requirement modeling, and valuable information generation. By introducing 5G messaging platform technology and utilizing the Rich Communication Suite (RCS) protocol, strong reach, lightweight interaction, rich media engagement, and intelligent services are provided. Through the synergy of 5G messaging platform technology and intelligent algorithms, personalized and high-quality weather services are realized, ultimately enhancing user experiences.

This research amalgamates user behavior analysis and data processing techniques, offering innovative approaches to realize personalized weather forecasting and smart meteorological services. It is poised to carve new pathways for refining the accuracy and personalization of meteorological services, delivering customized weather information and services to users.

2.1. Current State of Travel Prediction Research

Machine learning-based travel prediction methods are currently categorized into several types, including K-Nearest Neighbors (KNN) [13], Deep Neural Networks (DNN) [14], Random Forest (RF) [15], and the context-aware Loc-PredModel. KNN predicts by comparing similar travelers, DNN extracts traveler features through multi-layer neural network structures, while RF enhances prediction accuracy through ensemble learning.

Travel prediction, which utilizes historical or real-time data to forecast travelers' destinations, times, and other information, finds widespread applications in intelligent transportation, smart cities, the mobile internet [16,17], and personalized travel services [18]. Travel prediction methods can be broadly categorized into rule-based [19,20] and machine learning-based [21] approaches. Rule-based methods rely on manually set rules to infer travel patterns, but struggle with complex scenarios. In contrast, machine learning-based methods automatically learn travelers' characteristics and patterns through algorithms [22], making them suitable for high-dimensional and nonlinear data. In recent years, with the advancements in mobile internet, the Internet of Things (IoT), and big data technology, machine learning algorithms have achieved remarkable success in areas such as image recognition [23], speech recognition [24], and natural language processing [25], exhibiting great potential in the field of travel prediction. Among them, deep learning, through multi-layer neural network structures, can autonomously extract advanced features, showcasing significant promise.

This study aims to integrate travel behavior analysis and machine learning to achieve personalized weather forecasting based on user behavior. By extracting user travel data features and utilizing machine learning models, our goal is to enhance weather predictions to be more accurate and personalized. This approach holds the potential to open new pathways for improving the accuracy and personalization of meteorological services, expanding the prospects of weather prediction.

2.2. Current State of Personalized Weather Recommendation Research

Currently, research on personalized weather recommendations primarily focuses on user feature-based approaches. Studies aim to capture users' individual characteristics, preferences, and needs by constructing user profiles or models using historical behavior data or user feedback, thereby delivering personalized weather recommendations. For instance [26] utilizes association rules to recommend weather service products to users. Other studies propose recommendation methods based on user profiles and collaborative filtering. These methods analyze users' activities on social networks and location data to establish user profiles and employ collaborative filtering algorithms to suggest weather information that aligns with the interests of users similar to them.

Furthermore, some approaches integrate user models with context awareness, taking into account individual attributes such as health status and lifestyle habits, alongside the current context, to provide users with tailored weather information.

However, there is currently a lack of research on weather recommendations based on user travel behavior. Such research emphasizes the analysis of users' travel purposes, methods, times, and other behaviors. By utilizing historical or real-time travel data, these studies predict future locations or routes to offer users relevant weather information. Despite the existing progress in personalized recommendations based on user features, the field of personalized weather recommendations grounded in geographical location and travel prediction remains largely unexplored. Therefore, this study aims to address this research gap by comprehensively exploring and providing a travel behavior-based approach to personalized weather recommendations.
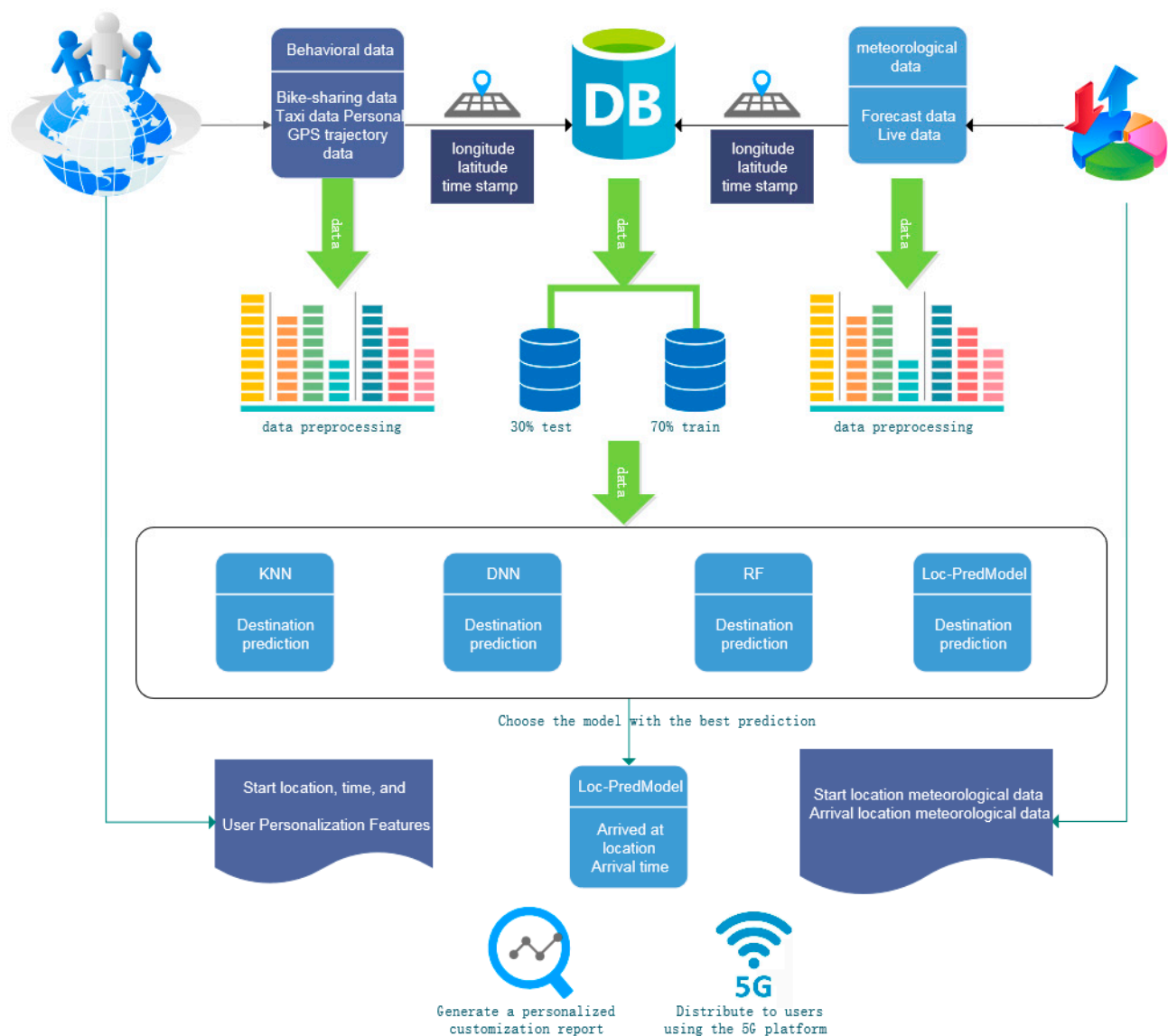
2.3. Research Content

The research framework of this paper aims to establish a personalized weather recommendation system based on travel and weather data. The framework primarily encompasses the following key steps, as depicted in the Figure 1:

Data collection and preprocessing: Initially, gather users' travel and weather data. Travel data encompass information such as users' travel purposes, methods, and times, while weather data include details like weather conditions, temperature, and humidity. These data need to undergo preprocessing to ensure accuracy and consistency.

Data association and fusion: Associate and merge travel data with weather data, connecting the two datasets using key information like latitude, longitude, and timestamps. This forms a comprehensive dataset, serving as the foundation for subsequent training and analysis of predictive models.

Destination prediction model: In this phase, compare and train four distinct prediction models to forecast users' travel destinations. These models might include the KNN model, DNN model, RF model, and our model. By assessing the performance of different models, select the most suitable one to accurately predict users' upcoming destinations.

Spatiotemporal weather report generation: Once the destination prediction model determines the upcoming user destination, combine spatiotemporal weather data for that location with user-defined preference features to generate a personalized customized weather report. This report not only includes standard weather information but also incorporates attire recommendations and activity suggestions based on user preferences.

Figure 1. Main research framework.

5G message push: Finally, utilizing the 5G message platform, precisely deliver the generated personalized weather report to users through push notifications. This platform enables high-speed delivery and rich media interaction, ensuring users promptly receive personalized weather information.

Through the aforementioned steps, the research framework of this paper establishes a personalized weather recommendation system based on travel and weather data. Consequently, it provides users with more considerate and accurate weather information services.

2.4. Innovative Aspects and Highlights

1. Fusion of travel behavior prediction and personalized weather services: This research introduces travel prediction techniques to the field of meteorological services for the first time. By analyzing users' travel behavior, it forecasts potential destinations and generates personalized weather forecasts for those destinations. This fusion combines travel behavior and weather data to provide users with more intelligent, real-time, and accurate weather services, thereby enhancing user satisfaction and experience.

2. Application of XGBoost algorithm in personalized recommendation model: This study employs the XGBoost algorithm, an efficient and powerful machine learning technique, to predict users' future locations. By leveraging user characteristics, this

machine learning model generates personalized weather reports that align with users' behavior patterns and preferences.

3. Construction of an intelligent weather push platform: By establishing a precise intelligent weather push platform, this research maximizes the broad coverage and multimedia transmission advantages of the 5G messaging platform. Integrating multiple data sources and predictive models, this platform delivers comprehensive, accurate, and personalized meteorological services to users. This push mechanism not only enhances the efficiency of conveying meteorological information but also provides users with smarter, personalized services.

4. Implementation of smart travel advice: The study further implements intelligent travel advice based on the 5G messaging platform, interacting with the intelligent weather push module to offer more accurate and timely travel recommendations. Combining machine learning techniques with spatiotemporal feature algorithms, this research accurately predicts urban travel status and trends, enabling users to receive more effective, granular weather reports, and enhancing their quality of life.

5. Real-world application and effect validation: This research goes beyond theoretical exploration and conducts practical trials in the Chengdu region, deploying the 5G-based intelligent personalized weather service platform. The successful implementation of this application demonstrates the feasibility and effectiveness of this approach, with the potential to elevate the quality of meteorological services and improve people's lives.

Through the aforementioned innovative aspects, this research presents novel ideas and opportunities for the application of travel prediction technology in meteorological services, the provision of personalized weather information, and the construction of intelligent push platforms.

In the subsequent chapters, the first section introduces the dataset used in the experiments, describing its source, collection method, and content. The second section outlines the methods and technologies employed in the research, delving into the machine learning-based travel prediction methods, including different types of models like KNN, DNN, and RF. It also details the principles and applications of the XGBoost algorithm, along with how user features are incorporated to generate personalized weather reports. Furthermore, the construction of the intelligent weather push platform is elaborated, including the utilization of the 5G messaging platform for precise delivery and the implementation of smart travel advice. The third section describes the experimental process and results, offering a comprehensive insight into the experimental design, process, and outcomes. It explains how user-specific features, travel patterns, and weather data were obtained from publicly available travel datasets and subjected to necessary preprocessing. The design and implementation of the machine learning models are meticulously discussed, along with how the preprocessed data were employed for model training and testing. In the concluding part, a summary of the entire research is provided. The experimental process and outcomes are revisited, analyzing the feasibility and effectiveness of the results. The research's innovative aspects and its role in addressing existing gaps are emphasized. Finally, limitations and future research directions are discussed to provide guidance for further investigations.

By comprehensively addressing these chapters, this paper offers a detailed exploration of various facets of the experiment, spanning from data processing to method application, experimental result presentation, and conclusion. This enables readers to gain an in-depth understanding of the research's contributions and value.

3. Dataset and Data Processing

This research conducts travel pattern analysis based on a shared bicycle dataset. Considering our focus on individual travel patterns, we opt to use transportation data as a proxy to indirectly extract individuals' spatiotemporal location patterns. In real-world scenarios, various modes of transportation, such as shared bicycles, taxis, buses, and subways, record features like departure time, departure location, arrival time, and arrival

location. Notably, shared bicycle datasets represent the travel patterns of a substantial number of urban users. Thus, we regard shared bicycle datasets as significant sources for studying individual travel patterns. Furthermore, 5G-enabled devices can also collect these data attributes. With this in mind, we adopt a transfer learning approach. We utilize publicly available shared bicycle datasets to construct travel prediction models and smoothly transfer these models to other modes of transportation and data collected from 5G-enabled devices for application.

Through the travel pattern prediction model, we can obtain information regarding users' arrival times and locations, which can be combined with user characteristics to offer personalized weather reports. To achieve this goal, this research necessitates the use of shared bicycle datasets in conjunction with corresponding weather datasets.

3.1. Bike-Sharing Data

The dataset employed in this study is derived from the Mobike bicycle dataset for August 2016 in Shanghai, which constitutes a shared bicycle open dataset. This dataset encompasses six distinct categories of features, namely bicycle ID (bike_id), user ID (user_id), departure time (start_time), departure longitude (start_lon) and latitude (start_lat), arrival time (end_time), and arrival longitude (end_lon) and latitude (end_lat). The dataset covers the timeframe from 1 August 2016, 00:00, to 1 September 2016, 00:00, and comprises a total of 1,023,603 records. Both Bicycle ID and user ID have been encoded using Label Encoder, while time data have been transformed into timestamps. Refer to Table 1 for the shared bicycle dataset fields and their corresponding descriptions. An illustrative example of the dataset is provided in Table 2.

Table 1. Shared bicycle dataset fields and their descriptions.

Field	Description
bike_id	Vehicle number
user_id	User ID, de-identified
start_time	Start time of the ride, accurate to the minute
start_lon	Start location (longitude)
start_lat	Start location (latitude)
end_time	End time of the ride, accurate to the minute
end_lon	End location (longitude)
end_lat	End location (latitude)

Table 2. Example of the bike-sharing dataset.

| orderid	bike_id	user_id	start_time	start_lon	start_lat	end_time	end_lon	end_lat
1	324731	7541	28 August 2016 0:00	121.49	31.28	28 August 2016 0:08	121.486	31.273
6	288841	4427	28 August 2016 0:00	121.462	31.313	28 August 2016 0:17	121.455	31.315
8	315873	6474	28 August 2016 0:00	121.416	31.154	28 August 2016 0:14	121.419	31.156
10	93155	12616	28 August 2016 0:00	121.436	31.32	28 August 2016 0:12	121.444	31.31
11	352484	11596	28 August 2016 0:00	121.459	31.325	28 August 2016 0:08	121.46	31.315
13	127840	9570	28 August 2016 0:00	121.353	31.285	28 August 2016 0:07	121.362	31.283
15	125990	15692	28 August 2016 0:00	121.441	31.228	28 August 2016 0:30	121.455	31.19
16	346549	4373	28 August 2016 0:00	121.426	31.222	28 August 2016 0:31	121.423	31.268
17	352552	350	28 August 2016 0:00	121.524	31.301	28 August 2016 0:14	121.518	31.291
20	351480	16317	28 August 2016 0:00	121.442	31.313	28 August 2016 0:10	121.453	31.311

The meanings of the fields in the data file are as follows, as indicated in the table below. Please note that some user information has been de-identified in the dataset.

3.2. Meteorological Observation Data

These weather data not only provide detailed information on weather conditions but also reflect the impact of different weather conditions on travel. Therefore, they become

key input factors in the intelligent weather push model. In this study, we also used meteorological observation data from the same time period as the Shanghai shared bicycle data, spanning from 1 August 2016, 0:00 to 1 September 2016, 0:00. These meteorological observation data meticulously records various weather elements during this period, such as temperature, humidity, wind speed, wind direction, precipitation, and more. By combining these meteorological observation data with the shared bicycle data, a more comprehensive description of weather conditions can be achieved for the study. Through the thorough utilization of these meteorological observation data, we can more accurately analyze travel patterns, thus achieving more precise personalized intelligent weather push services.

Figure 2 is the temperature chart for Shanghai in August 2016. Table 3. is example of Shanghai meteorological observation data.
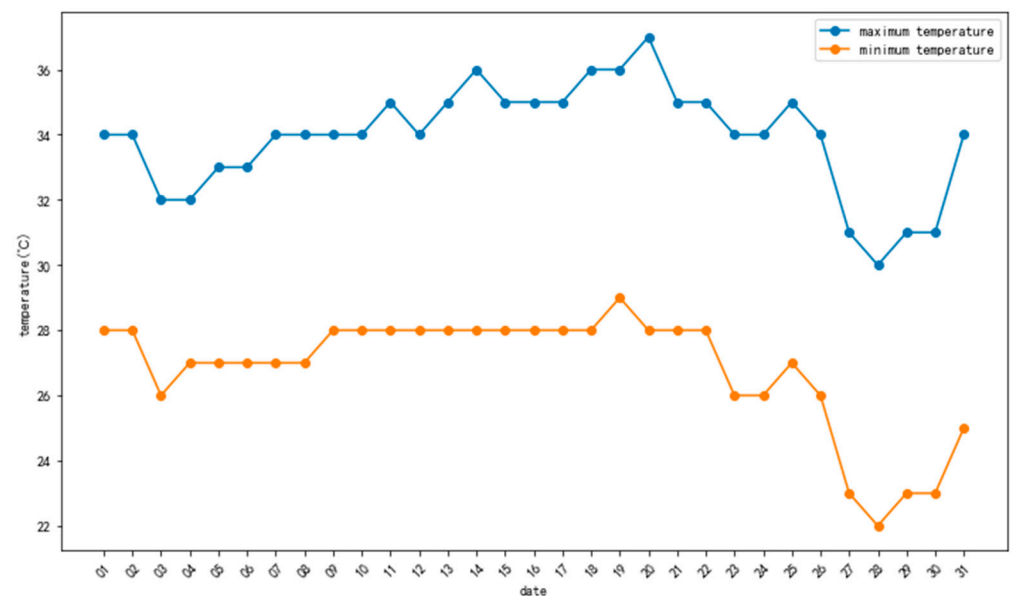
Figure 2. Temperature chart for Shanghai in August 2016.

Table 3. Table example of Shanghai meteorological observation data.

Datetime	Latitude	Longitude	Station Height	Temperature	Precipitation	Relative Humidity	Wind Speed	Wind Direction	Horizontal Visibility
03 August 2016 15:00:00	121.3667	31.1	5.5	27.4	0	81	1.5	119	32,500
03 August 2016 15:00:00	121.45	31.4	5.5	28.3	0	80	2.1	121	29,900
03 August 2016 15:00:00	121.1994	31.3806	6.5	28	0	79	1.5	161	35,000
03 August 2016 15:00:00	121.4928	31.6664	4.3	28.3	0	82	2.2	147	26,800
03 August 2016 15:00:00	121.4333	31.2	4.6	28.8	0	71	0.4	249	35,000
03 August 2016 15:00:00	121.7833	31.05	5	27	2.4	88	1.8	65	35,000
03 August 2016 15:00:00	121.5333	31.2333	4.4	28.8	0	75	0.9	99	27,300
03 August 2016 15:00:00	121.35	30.7333	5.2	28.8	0	71	3.4	44	35,000
03 August 2016 15:00:00	121.1167	31.1333	4	26.8	0.6	87	2	148	12,600
03 August 2016 15:00:00	121.2333	31.0333	4.2	28.3	0	73	2.6	88	35,000
03 August 2016 15:00:00	121.5	30.8833	4.6	28.8	0	75	3.5	75	35,000

3.3. Weather Forecast Data

Table 4 is example of Shanghai meteorological forecast data.

Table 4. Table example of Shanghai meteorological forecast data.

Datetime	Latitude	Longitude	Condition	Temperature	Ultraviolet Radiation	Visibility	Winddegrees	Wind Direction	Windlevel	Wind Speed
03 August 2016 15:00:00	31.231763	121.484443	Light Rain	28	1	700	270	West wind	2	2.69
03 August 2016 15:00:00	31.220367	121.424624	Heavy Rain	28	1	700	270	West wind	2	2.69
03 August 2016 15:00:00	31.223428	121.455965	Heavy Rain	28	1	1200	270	West wind	2	2.69
03 August 2016 15:00:00	31.24984	121.395555	Heavy Rain	28	1	700	270	West wind	2	2.69
03 August 2016 15:00:00	31.2646	121.505133	Moderate Rain	28	1	1200	270	West wind	2	2.69
03 August 2016 15:00:00	31.259541	121.526077	Light Rain	28	1	1200	270	West wind	2	2.69

3.4. Integration of Dataset Construction

In order to comprehensively analyze travel patterns and provide personalized intelligent weather recommendation services, this study adopts a method of associating shared bike data, meteorological observation data, and weather forecast data to construct an integrated dataset. This integrated dataset uses latitude, longitude, and time as reference points to organically combine data from different sources, providing richer and more accurate information.

Specifically, in the construction process of the integrated dataset, we first use latitude, longitude, and time as the basis for association, matching the departure and arrival locations and times from the shared bike data with corresponding timestamps in the meteorological observation data. This approach allows us to link each trip with the weather conditions at specific times, laying the foundation for subsequent analysis.

Simultaneously, we include weather forecast data in the integrated dataset to further enhance its richness and accuracy. Incorporating weather forecast data with actual observation data contributes to a more comprehensive weather prediction. By comparing forecasted data with actual observed data, we can better understand the changing trends in weather and provide users with more reliable weather information.

The construction of this integrated dataset allows us to comprehensively leverage multiple data sources, thus, more comprehensively capturing travel patterns and weather impact factors. By associating shared bike data, meteorological observation data, and weather forecast data, we can conduct more refined analyses in the spatial and temporal dimensions, providing stronger support for intelligent travel recommendations and personalized weather notifications.

3.5. Data Preprocessing

To effectively preprocess the shared bike and meteorological data, this study follows the following steps to ensure data quality and suitability:

1. Data cleaning: In the preliminary stage of data preprocessing, rigorous data cleaning is performed for both shared bike and meteorological data. Firstly, for shared bike data, records containing missing values or anomalies are removed. This includes data with unreasonable timestamps, latitude and longitude values exceeding actual ranges, and negative speed values. Similarly, for meteorological data, records with missing or abnormal values, such as invalid temperature, humidity, and wind speed data, are also eliminated.

2. Data transformation: After data cleaning, data transformation is carried out to suit subsequent analysis and modeling. Timestamps are converted into dates and hours, aiding in associating data with time to explore travel patterns and weather conditions across different time periods. Latitude and longitude values are transformed into grid

IDs, mapping spatial information to discrete grids for subsequent spatial analysis. Additionally, wind directions are converted into angle values to better comprehend and compare directional differences.

3. Data normalization: To eliminate scale differences between numeric data, normalization or standardization is applied. This ensures that data distributions are on a consistent scale, balancing the weights between different features. Normalizing data to a standard normal distribution or within the [0, 1] range prevents the model from being influenced by data scales during training and testing.

4. Data matching: In the process of merging shared bike and meteorological data, the key is to match the two types of data based on time and space. Each travel record is associated with weather observation data under a specific time interval and grid ID. This matching ensures that each trip is linked to weather conditions during its specific time period, providing strong support for subsequent analysis.

Through the aforementioned preprocessing steps, we obtain a normalized integrated dataset where each record includes shared bike data features such as user ID, time, latitude, and longitude, as well as meteorological data features like temperature, humidity, wind speed, wind direction, precipitation, and cloud cover.

During the experimental modeling process, data preprocessing plays a crucial role. Rigorous data cleaning, transformation, and matching guarantee the quality and accuracy of data used in model training and testing. Particularly with ample data available, effective data preprocessing is vital for establishing reliable analysis and prediction models. By removing problematic data records, we ensure the accuracy of model inputs, thereby enhancing the efficiency of predicting and analyzing travel patterns.

4. Methodology and Model Introduction

The methodology employed in this study comprises several key steps. Firstly, we collect and analyze users' travel behavior data to extract travel features and preferences. Subsequently, utilizing the XGBoost algorithm [27] from machine learning, we construct a predictive model for users' future locations, capable of accurately forecasting their upcoming destinations. Following this, personalized meteorological data reports are generated based on users' personal characteristics and the meteorological indicators of their higher interest. Finally, leveraging the 5G messaging platform technology, these personalized meteorological reports are precisely delivered to users, achieving an intelligent and personalized weather service.

4.1. Model Structure

Personalized weather push services can be customized based on predictable user travel information, such as where the user is likely to go and when they will arrive. In this paper, based on the Extreme Gradient Boosting (XGBoost) algorithm in machine learning technology, we build a user arrival destination prediction model, named Loc-PredModel, to integrates all kinds of data. The model takes the user's travel data (departure location, departure time, etc.) and additional data as input to the model to directly infer the destination location and arrival time of the user. The overall structure of the Loc-PredModel is shown in Figure 3.

The XGBoost algorithm is an improved algorithm based on the gradient-enhanced decision tree, adding additional features (such as column sampling and shrinking) to avoid overfitting and enhance the predictability of the model. By introducing a regularization term to measure the complexity of the tree model into the objective function, the risk of overfitting can be reduced. XGBoost can use a decision tree or a linear base model as the base learner, updating the weight of the learner based on the error obtained from each iteration. Finally, learners with different weights are combined to form an integrated model to implement the prediction. In general, the XGBoost algorithm tends to have better accuracy and less time to build the model due to its additional training process, while the algorithm also has the advantages of supporting parallel computation, built-in cross-

validation, and accepting missing values. Therefore, in this paper, the XGBoost algorithm is chosen to build the Loc-PredModel model for prediction. As an integrated tree model, the predicted value is calculated as shown in Equation (1):

$$\hat{y}_i^t = \hat{y}_i^{t-1} + f^t(x_i) \tag{1}$$

where $\hat{y}_i^t$ is the predicted value of the i th sample after the t th iteration, $\hat{y}_i^{t-1}$ represents the predicted value of the previous $t-1$ foundation model, and $f^t(x_i)$ represents the t th foundation model. Considering that the prediction of user arrival destination and arrival time is a regression problem, the loss function equation of the model is set as shown in Equation (2):

$$L = \sum_{i=1}^{m} l(y_i, \hat{y}_i) \tag{2}$$

where m is the number of samples, l is the training loss, y_i is the real value of the user's arrival destination and arrival time, and $\hat{y}_i$ is the predicted value of the destination and arrival time. After the regularization term is introduced, the objective function equation of the model is set as shown in Equation (3):

$$Obj = L + \sum_{i=1}^{n} \Omega(f_i) \tag{3}$$

This objective function consists of two parts, the former is a loss function and the latter is a regularization term, which is used to suppress the model complexity to prevent overfitting, where n is the number of trees. The expression of the regularization term is shown in Equation (4):

$$\Omega(f) = \gamma^T + \frac{1}{2}\lambda\|\omega\|^2 \tag{4}$$

where γ and λ are the penalty coefficients, T is the number of nodes in a given tree, and $\|\omega\|^2$ is the L2 regular term.
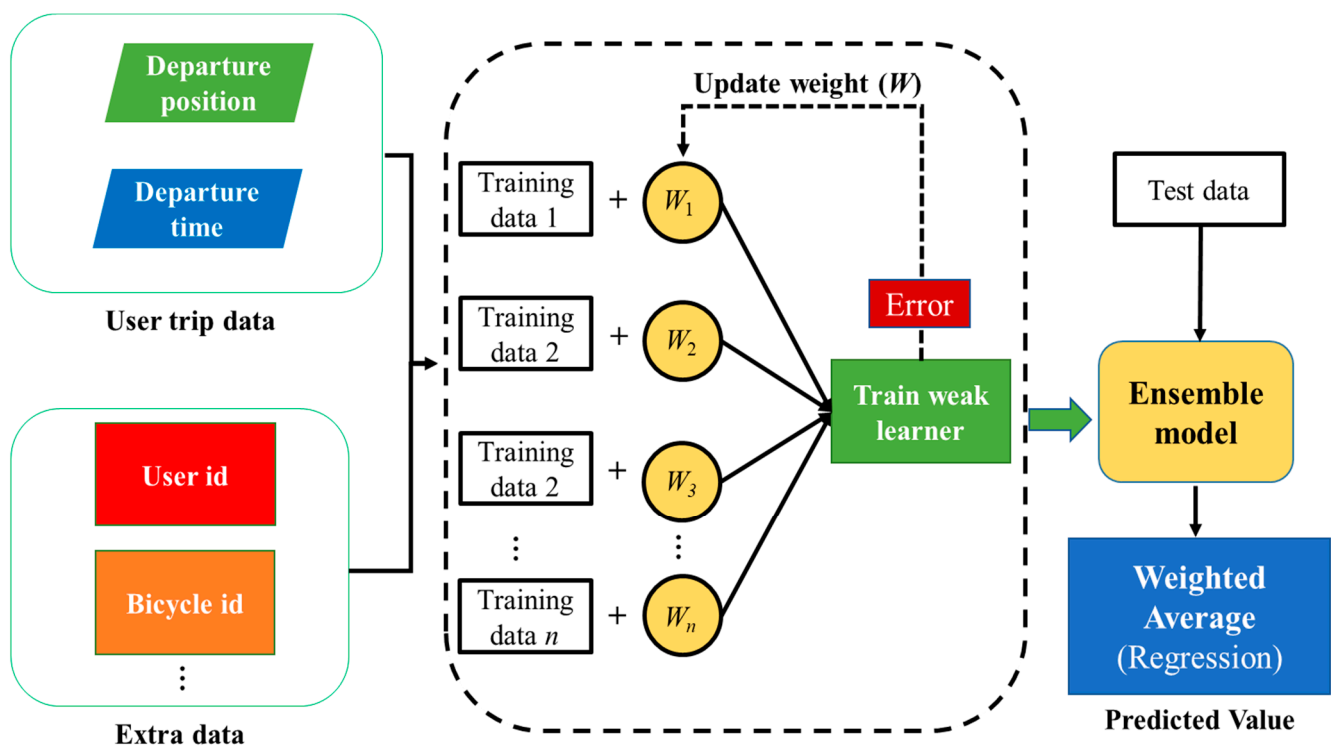

Figure 3. The structure of Loc-PredModel.

4.2. Model Hyperparameters

In this paper, a gridded parameter search method called GridSearchCV is used to determine the model hyperparameters. In the process of parameter search, the set of hyperparameters with the smallest error is selected as the final hyperparameters of the model. We conducted a grid search for six major hyperparameters of Loc-PredModel, and the search range of each hyperparameter is shown in Table 5.

Table 5. Hyperparameter search range in the GridSearchCV.

Hyperparameter	Range	Interval
n_estimators	100–600	100
max_depth	6–10	1
min_child_weight	2–4	1
learning_rate	0.1–0.5	0.1
subsample	0.7–1	0.1
colsample_bytree	0.7–1	0.1

5. Experiment and Evaluation

5.1. Experimental Environment and the Evaluation Index

The experiments were primarily conducted on a server with Graphic Processing Unit (GPU), with detailed software and hardware specifications outlined in Table 6.

Table 6. Experimental environment setting.

Item	Parameter
Operating system	Ubuntu 16.04
Memory capacity	128 GB
Central Processing Unit (CPU)	Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz
Graphics Processing Unit (GPU)	NVIDIA RTX 6000
Python version	3.8.2
Keras version	2.9.0

In this study, the performance of the models was evaluated using three metrics: mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and Pearson correlation coefficient (COR) [28]. MAE and RMSE were used to measure the deviation between predicted values and actual values, with smaller values indicating lower deviation. R^2 reflects the goodness of fit of the model to the data, ranging from 0 to 1, where values closer to 1 indicate a better fit. COR measures the correlation between predicted values and actual values, with higher values indicating stronger correlation. The formulas for these evaluation metrics are defined as follows, where n is the number of samples, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\overline{y}_i$ is the sample mean:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{7}$$

$$COR = \frac{COV(\hat{y}_i, y_i)}{\sqrt{Var[\hat{y}_i]Var[y_i]}} \tag{8}$$

5.2. Data Analysis Results

We begin by analyzing the features of user riding data in the dataset. Figure 4a presents the distribution of user riding distances. It is evident from the graph that the majority of users have riding distances not exceeding 2 km. Among these, approximately 42.4% of the data consist of distances under 1 km, while distances between 1 km and 2 km make up around 34.8% of the total data. Notably, data points with riding distances exceeding 5 km account for 10.8%, which could introduce higher prediction challenges due to the likelihood that longer distances reflect instances where users might change their riding destinations.

(a)

(b)
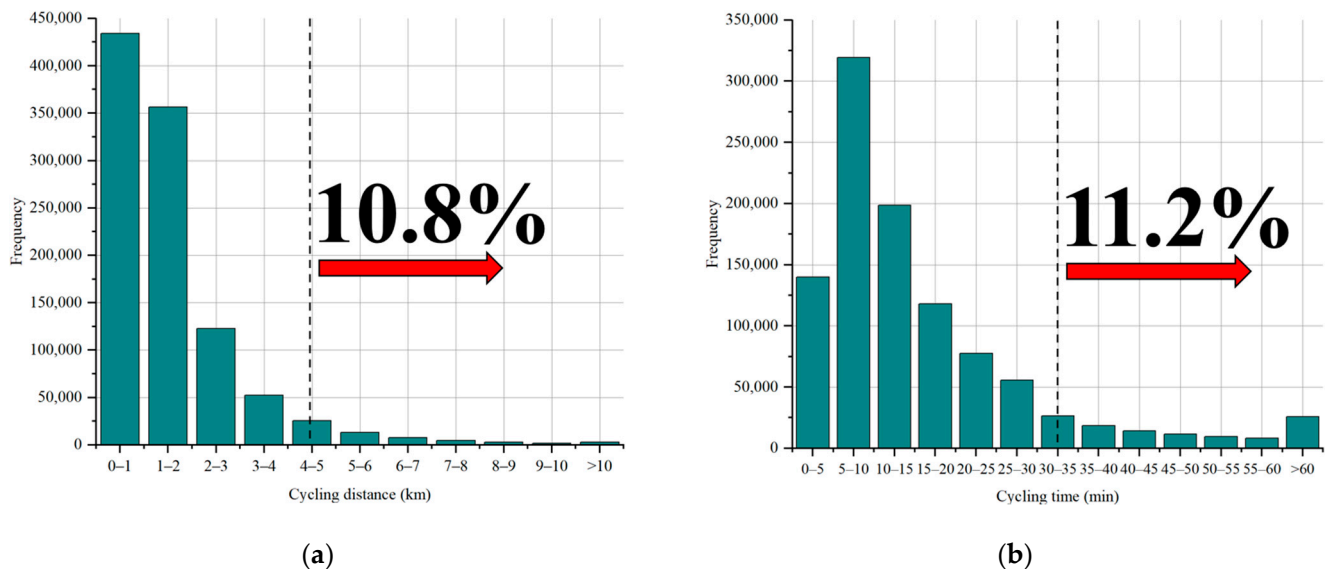
Figure 4. (**a**) Distribution characteristics of user cycling distance; (**b**) distribution characteristics of user cycling time.

Figure 4b portrays the distribution of user riding durations. The data show that the highest proportion falls within the range of 5 to 10 min, accounting for approximately 31.2%. Data points with riding durations exceeding 30 min constitute around 11.2%. Similar to riding distances, longer riding durations could also contribute to increased prediction complexities.

Next, we delve into the Pearson correlation analysis of various variables in the dataset. As shown in Figure 5a, the arrival time of users is highly correlated with their departure time, with a correlation coefficient of 0.94. Simultaneously, concerning users' arrival destinations, there exists a noticeable correlation with their departure locations. The correlation between arrival longitude and departure longitude is 0.68, while the correlation between arrival latitude and departure latitude is 0.88. Departure longitude also exhibits some influence on arrival latitude (Cor = 0.15).

Subsequently, using the XGBoost algorithm, this study conducts feature importance analysis on all input features. Figure 5b illustrates the results. It is evident that the departure time feature holds the greatest importance in the prediction process of the model, followed by the departure location's significance. Notably, the user ID and bike ID features have the least importance in the model's prediction, aligning with the results of Pearson correlation analysis among the variables.

(**a**)

(**b**)
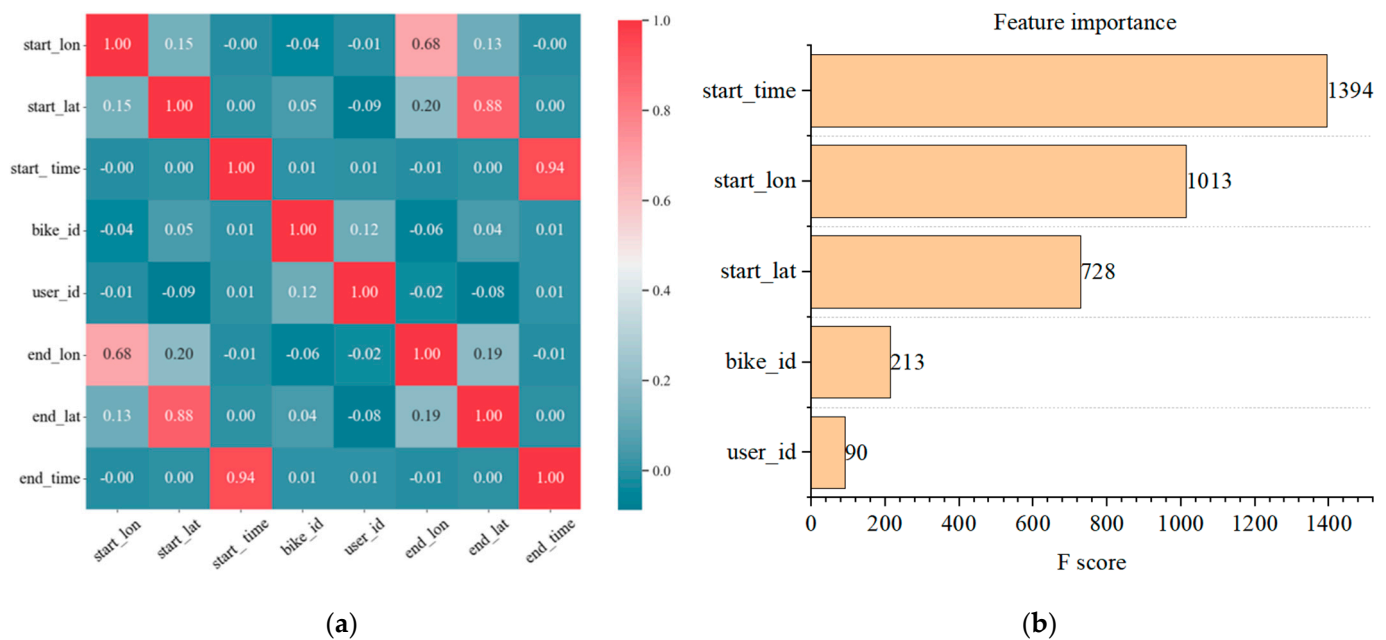
Figure 5. (**a**) Pearson correlation between variables; (**b**) feature importance of different variables from XGBoost algorithm.

5.3. Model Performance Comparison

The main features for predicting users' arrival locations and times include the user's departure location and time, with auxiliary features being the bike ID and user ID. The dataset was divided randomly into a training set (70%) and a test set (30%), which were preprocessed using standardization functions.

Several different models were developed and tested for predicting users' destinations and arrival times:

K-Nearest Neighbor (KNN): The KNN algorithm finds the K-Nearest Neighbors of a given sample in the feature space and assigns the average attributes of these neighbors to the sample for prediction. It was developed using the "sklearn" package in Python, with the parameter n_neighbors set to 7.

Deep Neural Network (DNN): The DNN receives predictive variables as inputs at the input layer, and generates predictions in the output layer by training neurons in hidden layers. It was developed using the "Keras" package in Python, comprising one input layer (128 neurons), two hidden layers (with 64 and 32 neurons, respectively), and one output layer, utilizing the "relu" activation function.

Random Forest (RF): This ensemble model is composed of multiple classification or regression trees. Input predictive factors are randomly partitioned into each tree using bootstrapping, training the predictive model using data within each tree to mitigate model overfitting risks. It was developed using the "sklearn" package in Python, with the parameter n_estimators set to 100.

Loc-PredModel: This model, designed for predicting users' destination using the XGBoost algorithm, was developed using the "XGBoost "package in Python. Parameters were set as n_estimators = 200 and max_depth = 3.

The performance of each model is presented in Table 7, where "Train_c" denotes the time consumption on the training set, and "Test_c" represents the time consumption on the test set. For location prediction, the units for mean absolute error (MAE) and root mean squared error (RMSE) are degrees, whereas for arrival time prediction, the units for MAE and RMSE are minutes.

Table 7. Model performance comparison.

Method	MAE	RMSE	R^2	COR	Train_c	Test_c
KNN	0.114	0.240	0.876	0.943	-	14.033 s
DNN	0.106	0.210	0.887	0.950	195.005 s	5.603 s
RF	0.105	0.248	0.911	0.953	16.216 s	1.029 s
Loc-PredModel	0.101	0.208	0.935	0.966	2.315 s	0.055 s

Overall, the Loc-PredModel outperforms other models across all three evaluation metrics. The performance of the Random Forest model is comparable to that of the Loc-PredModel. Notably, the Loc-PredModel exhibits the fastest efficiency, with a training time of 2.315 s on the training set and an inference time of 0.055 s on the test set. In contrast, the Random Forest model takes 16.216 s for training and 1.029 s for inference on the test set. The DNN model has the longest training time, at 195.005 s on the training set and 5.603 s on the test set. The parallel tree structure of the Loc-PredModel based on the XGBoost algorithm contributes to its lower time consumption, making it highly advantageous for practical deployment and application. The experimental results demonstrate the effectiveness of the Loc-PredModel in predicting users' arrival locations and times. It not only exhibits good prediction accuracy but also boasts lower time consumption.

After confirming the superior performance of the Loc-PredModel, further analysis was conducted, through feature ablation experiments, to explore the impact of different features on prediction performance. As shown in Table 8, the results show that the Loc-PredModel achieves good prediction performance when given the departure time and departure location information as inputs. Adding the bike_id feature has a larger impact on improving prediction performance compared to adding the user_id feature. This aligns with the results from the Pearson correlation analysis and feature importance analysis in Section 5.2. Importantly, when both bike_id and user_id features are included as inputs to the model, prediction performance does not improve further. This suggests that adding more features to the model does not always enhance prediction performance, as it might introduce redundant information.

Table 8. The impact of different predictors on model performance.

Features	MAE	RMSE	R^2	COR
start_time+start_lon+start_lat	0.101	0.268	0.908	0.950
start_time+start_lon+start_lat+user_id	0.101	0.218	0.928	0.963
start_time+start_lon+start_lat+bike_id	0.101	0.208	0.935	0.966
start_time+start_lon+start_lat+user_id+bike_id	0.101	0.208	0.932	0.965

Furthermore, we also examined how different travel distances and travel times affect the predictive performance of the model. This analysis helps us understand the model's ability to predict user behavior under various travel conditions. Figure 6a illustrates the predictive performance of the Loc-PredModel across different travel distances. Generally, the model exhibits better predictive performance for shorter travel distances and poorer performance for longer distances. The predictive ability of the model noticeably decreases when the travel distance exceeds 5 km (R^2 = 0.758). Figure 4b displays the model's predictive performance across different travel times. Similar to Figure 6a, longer travel distances correspond to poorer predictive performance. However, the decline in performance is more gradual, suggesting that the model is more sensitive to travel distance than travel time. These results confirm the hypotheses made in Section 5.2, which stated that the user's travel distance and travel time are significantly correlated with prediction difficulty.

(a)

(b)
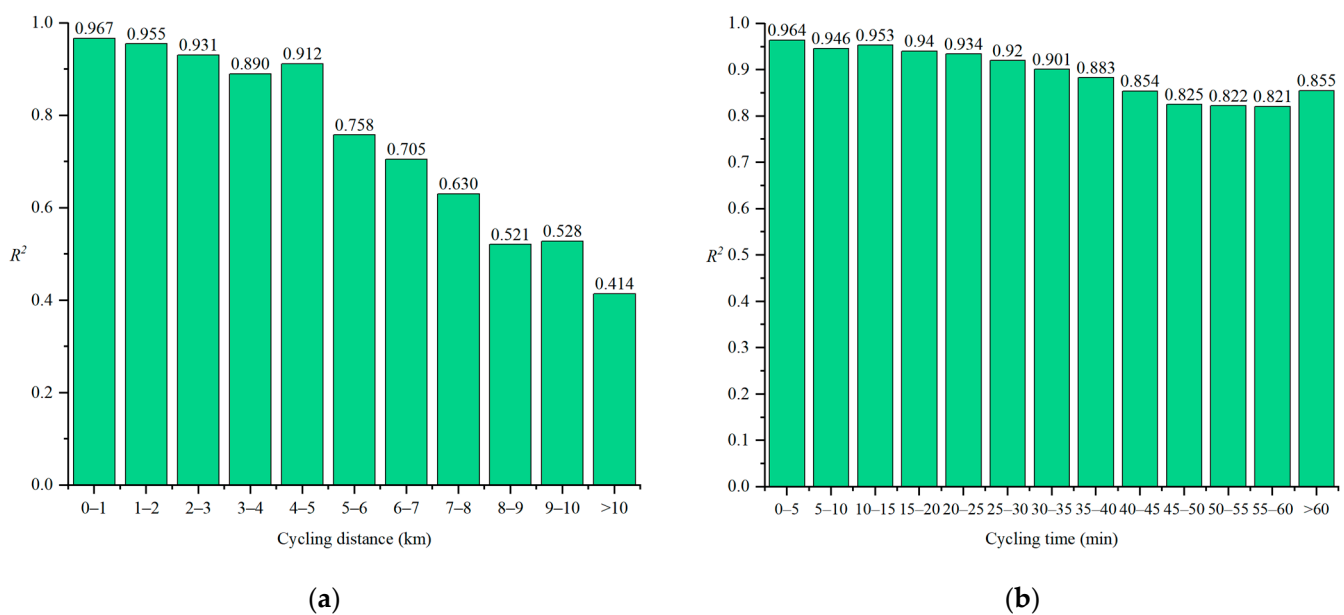
Figure 6. (**a**) The prediction performance for different cycling distances; (**b**) the prediction performance for different cycling time.

6. Personalized Weather Report Generation

The primary objective of this research is to deliver personalized weather reports that specifically address individual users' requirements and preferences. This section aims to provide a comprehensive understanding of how the outcomes from the predictive model are integrated with user preference features to formulate customized weather reports.

6.1. Consideration of User Preference Features

In order to generate personalized weather reports, it is crucial to take into account various user preference features, including their sensitivities towards elements such as temperature, precipitation, and wind speed. These features can be derived from historical preference data, user feedback, and personal information. By integrating these features, a better comprehension of user needs can be attained, thereby facilitating the creation of weather reports that align closely with their preferences.

6.2. Data Integration Process

The integration process involves the combination of three sets of data. Firstly, meteorological data comprise two categories: basic meteorological elements such as precipitation, temperature, wind speed, wind direction, and humidity, and lifestyle weather elements like air quality, UV index, dress index, and car wash index. Secondly, user travel data primarily include the user's mode of travel, departure time, departure location, arrival time, and destination. Lastly, user preference feature data involve user characteristics such as age group, occupation, and primary commuting method.

6.3. Integration of User Preference Features and Weather Data

After acquiring user preference features, they are integrated with the weather data produced by the prediction model to generate personalized weather reports. The process, illustrated in Figure 7, entails the recommendation system gathering weather data, user travel data, and user-specific features. Leveraging the Loc-PredModel previously discussed, the system predicts the user's location and time. Subsequently, the weather data, travel data, and user-specific feature data are transmitted to the application server, where they are amalgamated to generate data reports incorporating weather conditions, destination

location, and time. These personalized reports are subsequently disseminated to user terminals through the 5G messaging platform.
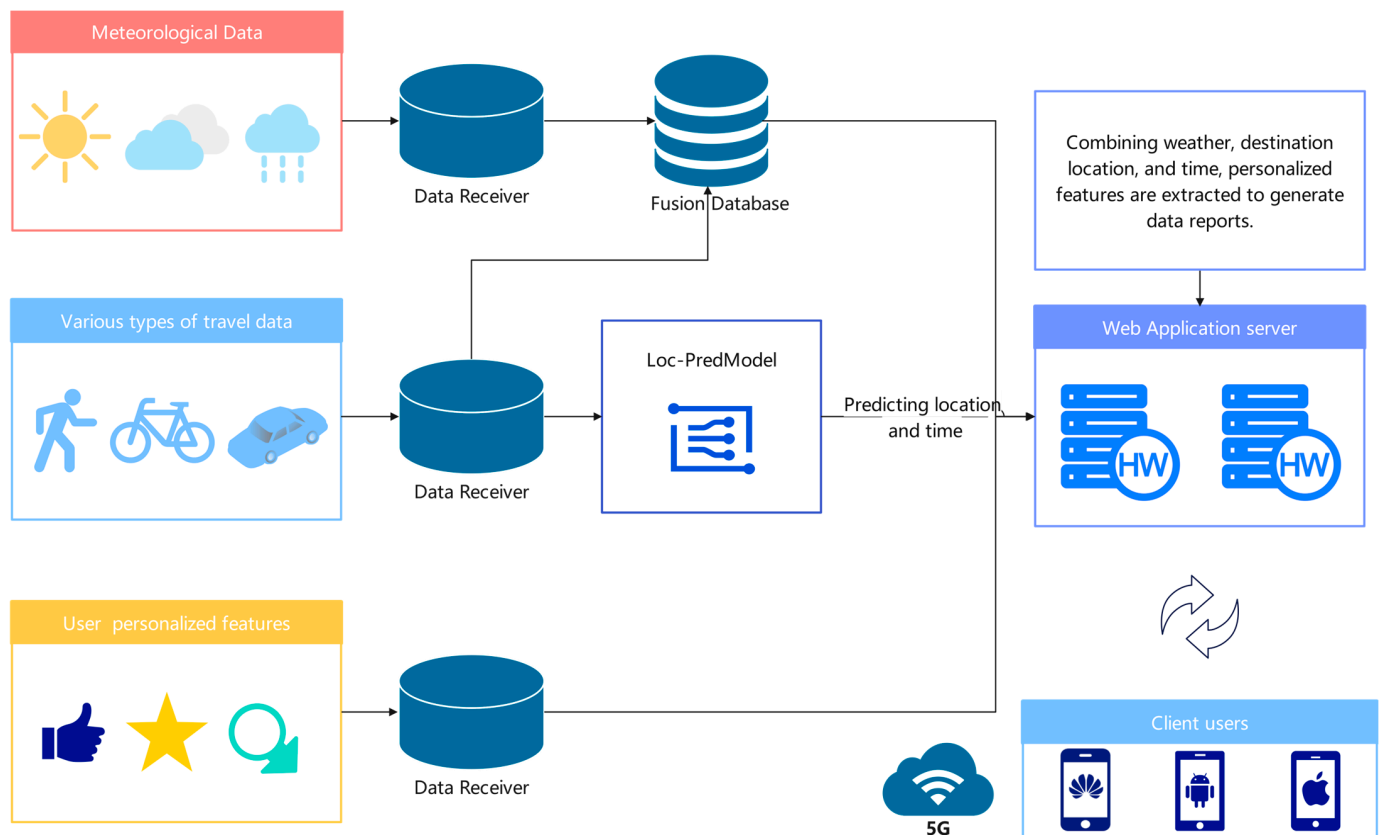
Figure 7. Architecture diagram of the intelligent weather push platform.

6.4. Practical Application Scenarios

As shown in Figure 8, 5G Messaging Cloud Service Platform, including functions such as notification management, reply configuration, 5G menu configuration, transmission management, rich media message editing, and data management, we have successfully implemented a personalized 5G Messaging Cloud Service Platform based on the Loc-PredModel. Our personalized weather service recommendation system aims to provide users with an intelligent, real-time, and tailored meteorological information experience. To achieve this goal, we have designed a variety of 5G message templates for different scenarios, catering to users' diverse weather information needs. These templates dynamically and automatically combine based on the user's current context and service variations, allowing flexible deployment of 5G message products for different scenarios. Whether users are on a journey, preparing for outdoor activities, or in need of real-time weather information, we can provide weather services relevant to their specific scenarios, ensuring users receive the latest and most accurate weather information.

In addition, our system implements user text recognition and multimedia weather interaction question–answer services across various scenarios. Users can pose weather-related questions to the system through text input or voice recognition. The system responds in rich multimedia formats, including voice synthesis, images, and videos, based on the user's question and current context. This enables users to intuitively grasp weather conditions, whether they are inquiring about today's weather, the temperature trend for the upcoming week, or seeking weather advice for specific activities. Our system caters to personalized needs, enhancing the user's meteorological service experience.
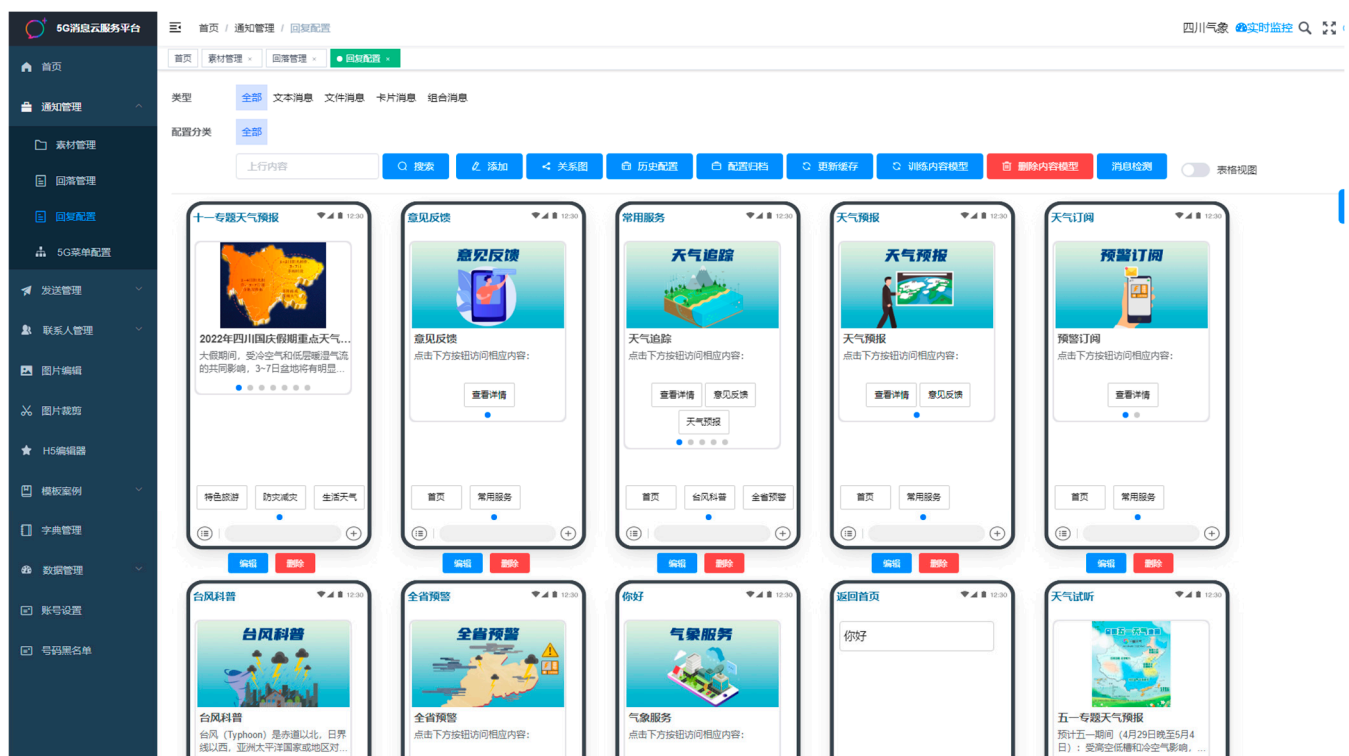
Figure 8. Deployed 5G Messaging Cloud Service Platform.

In summary, our personalized weather service recommendation system, through flexible 5G message templates and multimedia weather interaction question–answer services, aims to provide users with a more intelligent, real-time, and tailored meteorological service. We address users' diverse meteorological information needs, offering a more convenient and valuable weather forecasting experience.

7. Conclusions and Future Directions

In this research, a comprehensive exploration was conducted in the field of personalized weather recommendations to achieve accurate weather predictions based on user behavior. By analyzing various features such as user departure location, departure time, bike ID, and user ID, along with the correlations between weather information, predictions were made for user destination and arrival time, providing users with more personalized and practical weather recommendations.

Different model approaches, including K-Nearest Neighbor (KNN), Deep Neural Network (DNN), Random Forest (RF), and Loc-PredModel developed using the XGBoost algorithm, were thoroughly compared and analyzed. The experimental results demonstrated that Loc-PredModel performed exceptionally well across multiple evaluation metrics, particularly in terms of MAE, RMSE, R^2, and COR, showcasing its effectiveness in predicting user destination and arrival time. Furthermore, the Loc-PredModel exhibited significant advantages in terms of time efficiency, making it suitable for practical deployment and application.

It is worth noting that the feature ablation experiment further verified the impact of different features on prediction performance. It was observed that even with only departure time and departure location information as input, Loc-PredModel achieved satisfactory prediction performance. Moreover, incorporating the bike ID feature significantly improved prediction performance, while the improvement from the user ID feature was limited. Interestingly, the performance stopped improving when both bike ID and user ID features were simultaneously input, emphasizing the importance of feature selection and model design in personalized weather prediction.

In summary, this research delved into the relationship between user behavior and weather, resulting in a personalized weather prediction model based on user travel patterns. Experimental results illustrated the model's advantages in accuracy, practicality, and time efficiency. This not only offers a new approach to enhancing the accuracy and personalization of meteorological services but also provides robust support for the further development and application of weather prediction. With the continuous advancement of mobile internet and big data technology, we believe personalized weather prediction will play an increasingly crucial role in areas such as smart cities and intelligent transportation, providing users with more convenient and accurate weather information services.

By integrating 5G messaging platform technology and machine learning algorithms, this research proposed an intelligent weather notification model based on travel prediction and the 5G messaging platform. The model predicted users' future locations by analyzing their travel behavior patterns, combined user features to extract meteorological information they are most likely to be interested in, and generated personalized weather reports. These reports were then sent to users through the 5G messaging platform, supporting intelligent interaction between users and the server. This model realizes personalized smart weather services, enhancing user satisfaction and reliance on weather information.

8. Discussion and Limitations

The predictive capabilities of our location and time prediction model were initially limited by the constraints of the shared bicycle dataset, particularly in relation to temporal and spatial scales. However, the integration of spatial–temporal data obtained from 5G terminal feedback holds promise in overcoming these limitations. This incorporation enables our personalized weather report generation model to dynamically adapt to corresponding temporal and spatial constraints, facilitating the generation of tailored weather reports for effective dissemination to users.

8.1. Enhanced Predictive Capabilities

The implementation of 5G terminal feedback significantly enhances the acquisition of real-time spatial and temporal location data, thereby augmenting the predictive capabilities of our personalized weather report generation model. This real-time feedback mechanism not only enriches the dataset but also contributes to the refinement of the predictive model, leading to improved accuracy and relevance in the dissemination of weather reports to users.

8.2. Privacy and Generalizability Considerations

While the integration of 5G terminal feedback data provides promising opportunities, it necessitates robust data governance and privacy protection protocols. The ethical and secure utilization of user location data demands careful consideration and implementation of stringent privacy measures, including anonymization, data encryption, and transparent data governance. These measures are crucial in upholding user trust and data security, ensuring responsible and ethical deployment of the proposed model.

Moreover, the localized application of our approach within a specific geographical context may pose challenges related to the generalizability and adaptability of the model for diverse urban and rural settings. Variations in user travel behavior, environmental factors, and infrastructural disparities across regions call for comprehensive validation procedures to assess the model's efficacy in different geographic and climatic conditions. Comprehensive studies across varied geographical regions and user demographics are necessary to ensure the model's reliability and applicability across diverse settings.

Author Contributions: Conceptualization and validation, Y.Y.; formal analysis, writing—original draft and methodology, F.F.; data processing and revision, Y.L.; writing-review and editing, Y.X.; data curation, L.W.; formal analysis, H.Z.; software design, W.Y. All authors have read and agreed to the published version of the manuscript.

References

1. Li, N.; Fan, J. On the Importance of Meteorology in People's Lives. *Agric. Technol. Serv.* **2016**, *33*, 147–187.
2. Ma, L.; Fu, X. Theory and Application of Agricultural-Energy-Meteorology Coupling in Agricultural Energy Internet. *China Electr. Power* **2021**, *54*, 115–124.
3. Cao, R. Design and Implementation of Meteorological Service Products for Travel to and from School Based on WeChat Mini Program. *China New Commun.* **2021**, *23*, 30–31.
4. Shao, Y.; Qin, Z.; Li, X. Quality Control of Automatic Station Temperature Observations with High Temporal and Spatial Resolution Based on EOF. *J. Atmos. Sci.* **2022**, *45*, 603–615. [CrossRef]
5. Du, T. Discussion on Meteorological Data Mining Based on Cloud Computing and Association Rules Mining Technology. *Smart City* **2016**, *2*, 34. [CrossRef]
6. Dai, B.; Xu, Y.; Luo, R. Factors Influencing Information Overload of Social Media Users and Its Consequences. *Mod. Inf.* **2020**, *40*, 152–158.
7. Chen, L.; Han, B.; Li, B.; Li, S.; Wang, Y.; Ding, X. Analysis of Agricultural Meteorological Disaster Service Demand in Heilongjiang Province. *Disaster Sci.* **2019**, *34*, 78–82.
8. Zhang, X.; Han, H. Exploration of Personalized Service in Meteorological Research Institutions' Information Room. *Mod. Inf.* **2010**, *30*, 109–110.
9. Chen, X. Research on Improving Public Meteorological Service Capability. *Shandong Soc. Sci.* **2015**, *S2*, 159–160. [CrossRef]
10. Cui, X.; Guo, X.; Tang, J.; Xu, J.; Shen, M. Exploration and Practice of Personalized Agricultural Meteorological Service Based on SMS. *Hubei Agric. Sci.* **2012**, *51*, 4506–4509. [CrossRef]
11. Zhuang, G. Better Service Guarantee for the Modernization of a Strong Country through the High-Quality Development of the Meteorological Industry. *Study Times August* **2021**, *4*, 1.
12. *Circular of the State Council of the People's Republic of China on Printing and Distributing the Outline of High-Quality Development of Meteorology (2022–2035)*; Gazette of the State Council of the People's Republic of China: Beijing, China, 2022; pp. 11–16.
13. Keller, J.M.; Gray, M.R.; Givens, J.A. A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern.* **1985**, *SMC-15*, 580–585. [CrossRef]
14. Wen, W.; Wu, C.; Wang, Y.; Chen, Y.; Li, H. Learning structured sparsity in deep neural networks. In Proceedings of the Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29.
15. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning: Methods and Applications*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 157–175.
16. Liu, C. Analysis and Forecast of Traffic State in Chengdu Based on Taxi GPS Data. Master's Thesis, China University of Mining and Technology, Jiangsu, China, 2020.
17. Chen, H. Research on Temporal and Spatial Laws of Residents' Travel Based on Taxi GPS Data. Master's Thesis, Yunnan University, Yunnan, China, 2016.
18. Du, M. User Behavior Analysis and Demand Forecasting Based on Shared Bicycle Travel Data. Master's Thesis, Chang'an University, Chang'an, China, 2021.
19. Shi, W.; Gao, T.; Zeng, Y.; Liang, J.; Liu, C.; Cai, Y. Research on Travel Reminder Rules Based on Comprehensive Meteorological Analysis of the Whole Journey. *Guangdong Meteorol.* **2020**, *42*, 59–62.
20. Li, H.; Cao, M.; Wen, F. Construction of Intelligent Meteorological Service Model Based on User Behavior Analysis of Shenzhen Weather Mobile Internet Channel. *Prog. Meteorol. Sci. Technol.* **2019**, *9*, 222–224.
21. Li, B. Research on Destination Prediction Methods Based on Taxi Mobile Trajectories. Master's Thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2021.
22. Abdollahi, M.; Khaleghi, T.; Yang, K. An integrated feature learning approach using deep learning for travel time prediction. *Expert Syst. Appl.* **2020**, *139*, 112864. [CrossRef]
23. Zhang, Q.; Fu, F.; Tian, R. A deep learning and image-based model for air quality estimation. *Sci. Total Environ.* **2020**, *724*, 138178. [CrossRef] [PubMed]

24. Reddy, D.R. Speech recognition by machine: A review. *Proc. IEEE* **1976**, *64*, 501–531. [CrossRef]
25. Chowdhary, K.R.; Chowdhary, K.R. Natural language processing. In *Fundamentals of Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 603–649.
26. Zhao, W.; Liu, Y.; Yu, D. Intelligent Recommendation of Meteorological Services Based on Association Rules. *Big Data* **2018**, *4*, 72–85.
27. Chen, T.; Guestrin, C. In Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
28. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.