JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Optimising Bike Sharing Operations: Predicting Demand At the Station Level

1stMrs.K.Ravali
Assistant Professor
ComputerScience & Engineering
Vignan Institute of Technology and Science
Hyderabad,TelanganaState,India,

2ndD.Ramyasri
Student
Computer Science & Engineering
Vignan Institute of Technology and Science
Hyderabad,TelanganaState, India

3rdB.Navya
Student
Computer Science & Engineering
Vignan Institute of Technology and Science
Hyderabad,TelanganaState,India

4thYashwanthHatkar
Student
Computer Science & EngineeringVignanInstituteofTechnologyandScience
Hyderabad, Telangana State, India

***Abstract*— Bike sharing systems have evolved as critical compo-nentsofurbantransportationinfrastructure,providingsustainablean dconvenienttransitoptions.Theproperrunningofthesesystemsis strongly reliant on accurate demand forecasting, which is still adifficult issue due to the dynamic nature of user behaviour and ex-ternal factors such as weather and seasonal variations. Despitemuch study on demand prediction for bike sharing systems, thereisstillaneedtoaddressthecomplexityofstation-leveldemandpre-diction.**

Inordertoovercomethedifficultiesinforecastingbikerental demand at the station level, we present a thorough machinelearning-basedmethodinthiswork.Tocreateaccuratedemandes-timates, we blend historical rental data with real-time weather in-formation,seasonal trends,andanotherimportantformat.

Our approach attempts to optimise resource allocation forbike sharing operators while also improving user experience by in-corporating advanced machine learning methods such as DecisionTrees,RandomForest, and XGBoost.

The suggested system is implemented as a Flask web application,whichprovidesuserswithaneasy-to-useinterfaceforenteringdataandreceivingaccuratedemandprojectio nsatbothdailyandhourlyintervals. Through thorough testing and assessment, we show thatour technique improves the operational efficiency of bike sharingsystems. Our findings show considerable increases in demand pre-dictionaccuracy,resultinginhighersystemutilisationandusersat-isfaction.

This study advances the administration of bike sharing sys-temsbyprovidingaworkableandeffectivemethodforanticipatingdem andatthestationlevel.Byprovidingoperatorswithactionableinsightsi ntopredicteddemandpatterns,oursolutionallowsforim-provedresourceallocation,loweroperationalcosts,andasmootheruse rexperienceinbikesharingsystems.

Keywords*— *Bike Sharing, Demand Prediction,

MachineLearning,Station-

Level,Optimization,FlaskWebApplication

I. INTRODUCTION

1.1 MotivationfortheProblem

Thefastdevelopmentofbikesharingsystemsinurbanarea shastransformedthetransportationlandscape,providinga moresustainableandenvironmentallyfriendlyalternativet otraditionalmodesoftransit.Thesesystemsallowcustomer stoaccessbikesfromseveralstationsthroughoutthecity,enc ouraginglast-milecon nectionandloweringcarbonfootprints.However,theabilit y to accurately estimate and manage user demandisimportanttothesuccessof bikeshare programmes.

Accurate demand forecasting is critical for bike shar-ing operators to optimise resource allocation, maintainsufficient bike availability at high-demand stations, andimprove overall system performance. The dynamic na-ture of user behaviour, combined with external factorssuch as weather, holidays, and seasonal variations, cre-ates substantial hurdles in constructing exact demandprediction models. Despite advances in machine learn-ing and data analytics, there is an urgent need to tacklethesedifficulties atthe station level.

1.2 IntroducingUnknownAspects

The complexities of station-level demand forecastingin Bicycle sharing systems add a number of unknownvariables that necessitate sophisticated modelling tech-niques. Factors such as changing consumer preferences,temporaltrends,andtheimpactofweatheronbi kerentalpatterns all pose important challenges. Understandingand capturing these distinctions is critical for construct-ingreliablepredictionmodelscapableofaccuratelyforecast ingdemand atspecificstations.

1.3 ContributionsoftheProposedSystem

In response to these problems, we introduce a novel ma-chinelearning-basedapproachdesignedexclusivelyforsta-tion-level demand forecast in bike sharing systems. Oursuggested system generates accurate and trustworthy de-mand estimates by combining several data sources such ashistorical rental data, real-time weather information, holi-day schedules,andseasonaltrends.

Oursuggestedsystemmakesthefollowingmajorcontribu-tions:
1. Advanced Machine Learning Models:Using cutting-edge algorithms like Decision Trees, Random Forest, andXGBoosttoidentifycomplicatedpatternsinbikerentalde-mand.
2. Integrated Data Fusion: Using several data sources toconstruct a demand prediction framework that considerstemporal,geographical, andexternal aspects.
3. Implementedauser-friendlyFlaskwebapplicationtoen-able bike sharing operators to simply input data and antici-patedemand overtime.

II. BACKGROUNDANDDATASET

2.1 BackgroundWork

Bikesharing systemshaveincreased in popularityincitiesaround the world, providing a flexible and environmentally ben-eficial means of transportation. Efficiently administering thesesystems necessitates precise forecasting of bike rental demand,whichisinfluencedamultitudeofconditions,includingweather, temporal trends, and user behaviour. In this portion, wewill present a synopsis of previous research in the topic of bikesharing demandprediction.

2.2 TraditionalApproaches

Earlyresearchonbikesharingdemandforecastfrequentlyused Classical statistical methods like time series analysisand regression models. Autoregressive Integrated MovingAverage(ARI-MA)andLinearRegressiontechniqueswereutilisedtodescribetemporalpatternsandlinearcorrelationsbetween demand and variables for example, time of day,day of week, and seasonality. While these methodologiesyieldedusefulinsights,theyfrequentlyfailedto capturethecomplex and non-linear correlations found in bike sharingdata.

2.3 EnsembleLearning

To overcome the limitations of traditional methodologies, re-searchers used ensemble learning methods such as Random For-estandGradientBoostingMachines(GBM).Ensembleap-proachesusemanybasemodelstoincreaseforecastaccuracyandrobustness. Studies have demonstrated that Random Forest, inparticular, is good at capturing the interactions between many fac-torsimpactingbikerentaldemand.Ensemblelearningtechniqueshave improved demand forecasting models' predictive power byusing the collective wisdomofmultiple models.

2.4 DeepLearningMethods

With the introduction of deep learning, researchers beganto investigate the use of neural networks in bike sharing de-mand prediction. Recurrent Neural Networks (RNNs) andLong Short-Term Memory (LSTM) networks have demon-stratedpotentialforcapturingtemporalconnectionsandlearning complicated patterns in sequential data. These deeplearning architectures excel at simulating the dynamic natureofbikerentaldemand,particularlywithlong-termdependen-ciesandnonlinear interactions.

2.5 FeatureEngineeringandDataFusion

Featureengineeringanddatafusionarecriticalcomponentsof accurate demand forecast in bike sharing systems. Re-searchers have worked on extracting significant elementsfrom a variety of sources, including weather data, calendarevents, geographic information, and station attributes. By in-cluding these factors into predictive models, researchers en-hanced the accuracy and granularity of demand estimates.Data fusion approaches, which incorporate information fromnumerous datasets, have also helped to improve model pre-diction skills.

2.6 ChallengesandOpportunities

Despite advances in demand forecast techniques, signifi-cantobstaclesremaininthefieldofbikesharingsystems.Thedynamicaspectofuserbehaviour,Theeffectsofexternalfac-tors such as weather and events, and the spatial-temporal in-tricaciesof bikesharingnetworksall providecontinuingstudyopportunities.Futureresearchmayconcentrateoncon-structingadaptivemodelsthatcanalterforecastsinrealtime,investigating the usage of sophisticated deep learning archi-tectures, and integrating contextual information to improveforecasting accuracy.

DATASET

Thedatausedinthisstudycamefromabigcity'sbikeshar-ingscheme.Itprovideshistoricalrecordsofbikerentals,withthefollowingessentialfeatures:
DateandTime:Thetimestampforeachbikerental.Holiday:Abinaryvariablethatindicatesifthedayisaholi-day(1or0).

```
def timeOfDay(hour):
  if hour == 17 or hour == 18  or hour == 19 or hour == 16:
    return 5
  elif hour == 15 or hour == 11 or hour == 13 or hour == 12:
    return 4
  elif hour == 14 or hour == 7 or hour == 9 or hour == 8:
    return 3
  elif hour == 10 or hour == 20 or hour == 21 or hour == 6:
    return 2
  elif hour == 22 or hour == 23 or hour == 0 or hour == 5:
    return 1
  elif hour == 1 or hour == 2 or hour == 4 or hour == 3:
    return 0
```

Figure1. Showshowtripsareclassified by timeofday.

Table1: Datasetdescription.

Name ofDataset	no.ofRows	no.of Columns
Data1	100	5
Data2	200	10
Data3	300	15

Weather Situation: A categorical variable describing the currentweather conditions. The bike sharing demand prediction systemusesaFlaskwebapplicationastheuserinterface.Temperature isa numeric variable. Users can enter critical factors such as thedate, holiday status, weather conditions, temperature, humidity,and windspeed using this simple interface. This front-end inter-face simplifies the process for bike sharing operators, offeringthem easy accessto thesystem'spredictioncapabilities.

Machine learning models, trained on past bike rental data, are attheheartofthesystem.Thesemodelsincludedecisiontrees,ran-domforests,andgradientboostingalgorithms,allchosenfortheirability to reliably estimate bike rental demand. These algorithmscan make accurate predictions based on user input data by lever-aging patterns and insights acquired from previous rental behav-iour.

The system features a feature engineering module to help withprediction. This module is in charge of analysing the input dataandextractingkeyvariablessuchasseason,year,weekday,work-ingday, and weather conditions. By generating these featuresfrom user-provided data, the system ensures that machine learn-ing models receive the inputs they require to make accurate anddependablepredictions.

Whenauserprovidesinput,thesystemselectsandpredictsmod-els. The algorithm chooses the proper model based on whetherthe user wants to anticipate the entire day or a specific hour. Theinput data, together with the extracted features, are then fed intothe chosen model to predict. This procedure ensures that the sys-temproducesspecialisedandpreciseforecastsbasedontheuser'sspecificconditionsandvariables.

Overall,thesystem'sarchitectureisintendedtoseamlesslyinte-grate user interactions, machine learning modelling, feature en-gineering, and prediction procedures. By providing a stream-lined and intuitive interface, paired with sophisticated machinelearning algorithms, the system empowers bike sharing opera-tors to make data-driven judgements, optimise resource alloca-tion,alongwithimproveeffectivenessingeneral.

Link:https://www.kaggle.com/datasets/way2tutorials/bike-sharing-dataset

III. SYSTEMARCHITECTUREANDSIMULATIONS

The system architecture allows users to estimate station-specific demand using a React web application. Users interactusing browsers, with React communicating using the Flaskbackend. via RESTful APIs. Flask server queries a database,usesapickleserialisedmodelfordemandestimation,an dre-turns the results to React for presentation. Figure 2 depictsthe system's flow, which demonstrates how user interaction,database activities, and predictive modelling are seamlesslyintegrated for efficientdemandassessment.

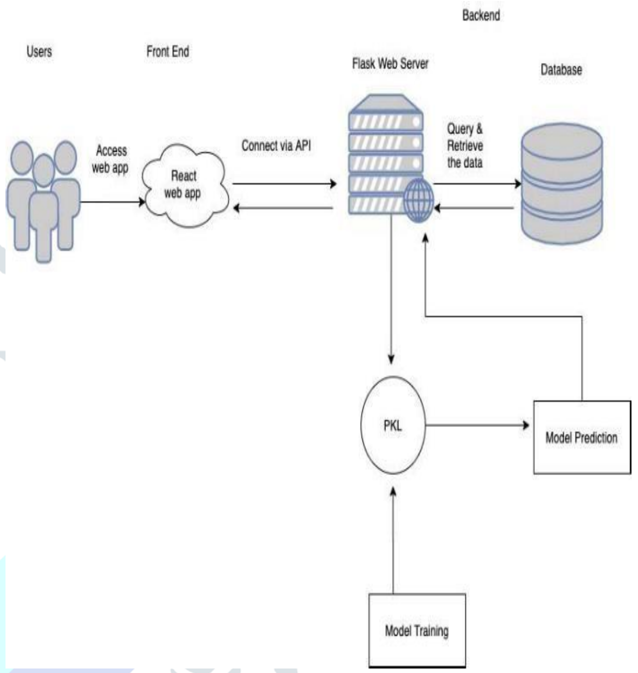


Figure2:StructureoftheBikeSharing ApplicationSystem

Users interact with the web application by visiting the homeendpoint,wheretheycaneasilyviewstationinformationsuchas ID, name, current bike demand, and availability status. Thesystem's model forecasts bike demand quickly, using datasuchasclusterID, Dayofweek,year toprovidepreciseesti-mates. Figure 3 depicts the complete process, from demandcomputationtouserdisplay,inanintuitive layout.

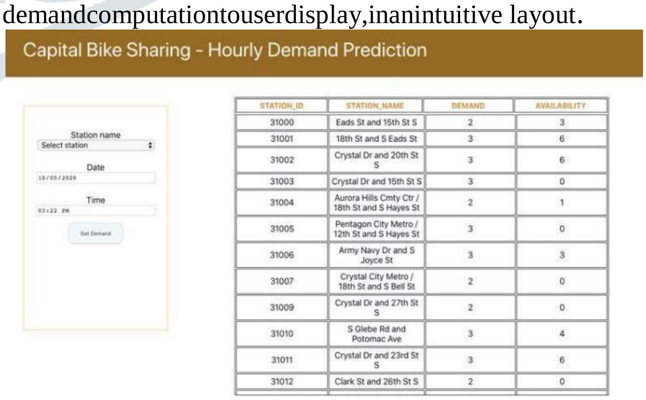


Figure 3: Website displaying the current demand and availability statusof all stations.

The user utilizes a dropdown menu to choose a name of the station in available record. After selecting a specific day and time,they proceed by clicking the "Get Demand" option. When thechosen a date is either in the past or future, the availability statusis marked as "Unknown." However, the bike demand is fore-casted by inputting cluster ID, season, weekday, and hour attrib-utes into the algorithm. The React web application then displaysthe station's details including ID, name, available bikes count,bike demand, along with those of the four related stations, as de-picted inFigure4.

InFigure4,thewebpagepresentsthecurrentdemandandavailabilityinformation forthestation selectedby theuser.

Our machine learning models, including Random Forest, Gradient Boosting, and Linear Regression, were constructed and comparedforperformance.Theoptimalmodelforprojectimplementationwasdeterminedtobe LinearRegressionusingtenclusters.This choiceresultedinaneffectivemodelduetothesimilarityinthe range of input values, resulting in a dense model. However,thedatasetwith20clustersledtoasparsemodelwithlessfavor-able outcomes.

Weemploythe classifierslisted below:

- Decision Trees: These are visual representations of potentialresponsesdeterminedbyspecificparameters.
- The process of bagging is a meta-algorithm for ensembles. that combines predictions from multiple decisiontreesusingOverwhelmingdecision.
- Random Forest:The Bagging based method that constructsacollectionofdecisiontreesbyrandomlyselect-ing a subset offeatures.

Random Forest, as a bagging-based technique, constructs an ensemble of decision trees by randomly selecting subsets of features,enhancingitspredictiveperformance.

The Random Forest method exhibited suboptimal performanceduetotherelativelylowcardinalityofdependentvariables. Despite randomly selecting features, the model resulted in Additional leaf nodes having dimensions 1 as wellas 2. The predictions typically ranged around 1 and 3, yield-ing an R-squared number is nearer to one, as illustrated inTable 2.

Dataset	Estimators	Maximumleaf nodes	Maximumdepth	Minimumsample splits	Resulting R2_Score
Data1	300	4	None	10	0.23467
Data2	1,000	5	10	30	0.42362
Data3	1,000	5	10	30	0.25055

Table 2 displays the random forest of parameters as well ascorrespondingaccuraciesforthewholedataset,alongsidesubsets containing10and20 clusters.

Figures 5 to 7 illustrate the outcomes of the Random Forestmodel across the three datasets. However, due to its limita-tions inhandling outliers effectively, the RandomForesttechnique exhibited subpar performance across all studiedbike-sharedatasets.

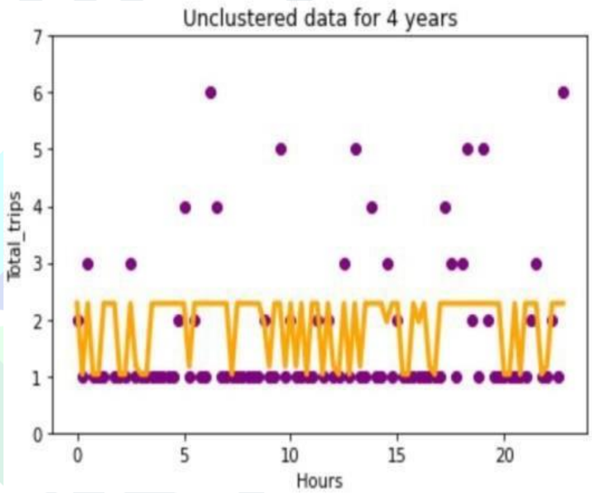
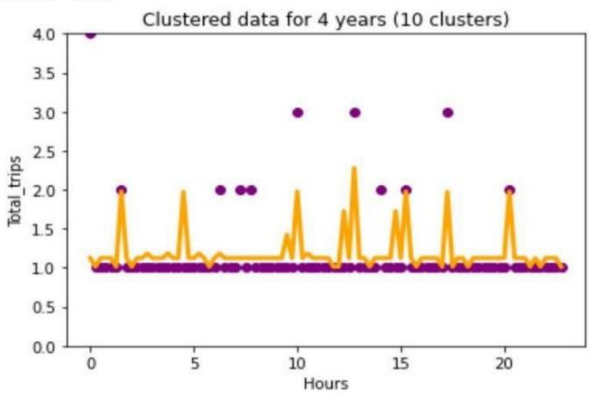


Figure5illustratesthecomparisonbetweenactualalsoforecasted data employing the random forest applied tounclustereddataoveraspanoffouryears.Theactualdatais depicted in purple, while the predicted values are represented inorange.



In Figure 6, the plot displays the comparison betweenactual and estimated information utilising the randomforest method with grouped information spanning fouryears (10 groups). The actual data is depicted via pur-ple, while the forecasted outcomes are represented inorange.

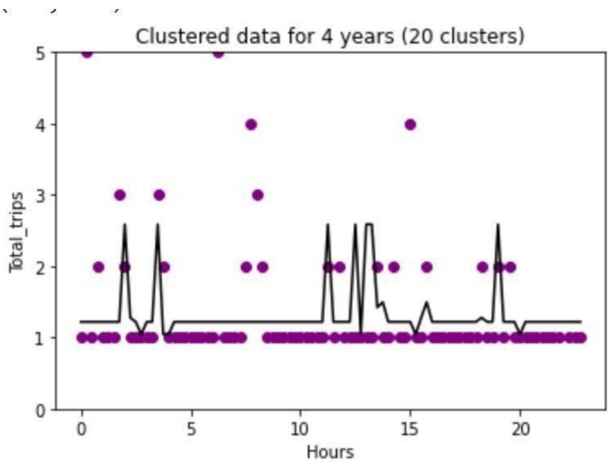


Figure7presentsacomparisonofactualandanticipateddatawiththe random forest method on grouped information spanning fouryears (20 groups). The actual data is shown via purple, while theestimated outcomesare depictedvia black.

Gradient Boosting as well as XGBoost progressively establishmodels, reducing errors from previous iterations and enhancingthe influence of top-performing models. Gradient Boosting uti-lizesthegradientdescenttechniquetoreduceerrorsinsucceedingmodels.Ontheotherhand,XGBoostisanimprovedslopeboost-ingmethodthatmitigatesoverfittingandbiasthroughtechniquessuch likeconcurrentprocessing,treetrimming,absentvaluehan-dling,andnormalization.

Adjustments to max_depth and R-squared settings have beenmadetoimprovepredictionaccuracyfornon-clustereddata.XGBoost, due to its ability to combine various weak learners ef-fectively, shows notable proficiency in learning station-specificdata patterns. In our analysis, XGBoost outperformed all othermodels on all datasets, leading to its selection for the web appli-cation'simplementation.

The successful execution of the algorithm heavily relies on pa-rameterchoices.Optimaloutcomescouldbeachievedbypickingsuitable algorithm variables based on application context. Fol-lowing a thorough evaluation of various settings, the most effec-tive values were selected. The XGBoost hyperparameters andcorresponding accuracies for bike-share datasets are detailed inTable 3.

Name ofDataset	Estima-tors	Maximumleafnodes	Maximumdepth	Minimumsamplesplits	Resulting R2_Score
Data1	300	17	None	20	0.7502
Data2	450	30	10	15	0.9304
Data3	300	15	10	10	0.9409

The XGBoost model consistently demonstrated strong perfor-mance across all datasets, achieving R-squared values close toone. The model was taught with a machine learning approach,as well as its was efficiency evaluated on training collection.Figures 8, 9, and 10 depict the favorable outcomes of theXGBoostmodelacrossdatasets.Additionally,theXGBoostal-gorithm exhibited robust performance across all datasets andeffectively managedoutliers.

Figure 8 shows a plot of true and forecasted data using theXGBoost algorithm on non-clustered data across four years.Truedataisgiveninpurple,whereasanticipatedfindingsareshown inorange.

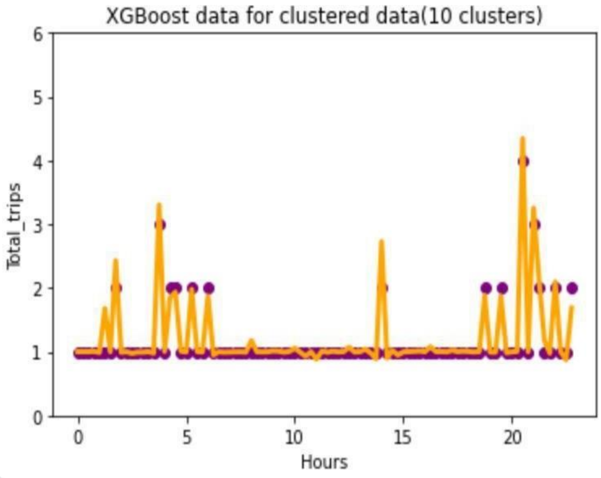
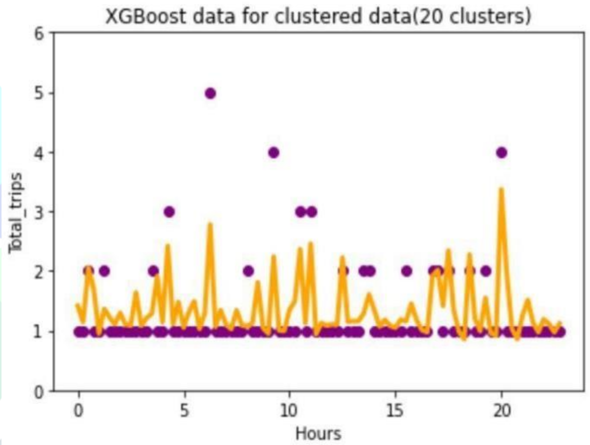


Figure 9 shows a plot of true and forecasted data from theXGBoost algorithm on clustered data across four years (10clusters). True data is given in purple, whereas anticipatedfindingsare showninorange.



In Figure 10, the plot illustrates the comparison betweenactualandestimateddatawiththeXGBoostappliedtogrouped data spanning four years (20 groups). The true datais represented via purple, while the anticipated outcomes aredepicted inorange.

Performance evaluation was conducted by K-fold verifica-tion, where the dataset is divided into k=10 equal-sized bins.Subsequently, k=10 distinct learning experiments are exe-cuted. In each run, one of the k subsets is designated as thetest set, while the remaining bins are amalgamated to formthe training set. The training set accuracy for K-fold cross-validation was determined to be 0.93, with the test set cor-rectness measuring at 0.83. This cross-validation procedurewas repeated 10 times, and the normal efficiency was com-puted acrossthetendifferenttestsets.

IV. CONCLUSION

Bicycle sharing demand prediction system represents asignificant advancement in urban transportation management, combining machine learning capabilities withdataanalytics.Itsprecisepredictivealgorithmspavetheway foroptimalbikelocation,improvingoperationalefficiencyforbikesharingoperators.Byutilisinginsightsonf actors

ThetemperatureisindegreesCelsius.

Humidity: A numerical quantity that represents the relative humidity. Windspeed: A numeric variable that indicatesthe wind.

The system provides operators with data-driven decision-makingcapabilitiesbasedonvariablessuchasweather, holidays, and historical demand trends. Thisnotonlyimprovescustomerpleasurebyincreasingbikeavailability,butitalsoresultsinreal costsavingsbyreducingresourcewaste.Finally,thesystemactsasacornerstone for sustainable urban mobility, giving a modelfor smarter, more efficient bike sharing ecosystems thatalign.

V. FUTURESCOPE

Movingforward,therearevariousoptionsforfurtherresearchanddevelopmenttofurtherstrengthenthebike-sharingdemandpredictionmechanism.

1. Integrationofreal-timedatasources,suchasliveweatherand traffic updates, can help enhance prediction accuracy.Thisallowsthealgorithmtochangepredictionsbasedoncur-rentenvironmentalandsituationalcharacteristics.
2. Analysinguserbehaviourdata,includingpreferredroutesand usage habits, can offer useful ideas on demand trends.Thisdatacanbeutilisedtocustomiseservices,improvebikedistribution,and customisepromotional campaigns.
3. AdvancedMachineLearningModels: Utilisingadvancedmachinelearningalgorithmsandapproachescanimprove prediction accuracy.
Deep learning and emulation strategies could increase modelperformance.
4. CollaborativePartnerships:Workingwithmunicipalplanners,transportationauthorities,andbikeadvocacygroupscan enhancethe system'sdatasourcesand insights.
Shared data projects can result in comprehensive city-widebikesharingsolutionsandencouragesustainableurbantrans-portationona largerscale.

By addressing these future work areas, the bike sharing demandforecastsystemwillbeabletoevolveandadapt,maintaining its relevance and effectiveness in optimising bikesharing operations and promoting sustainable urban trans-portation

REFERENCES

[1] N.Nibras.PredictingnumberofBike-ShareUsers.PublishedNov.2018.Available onlineathttps://towardsdatascience.com/predicting-no-of-bike-share-users-machine-learning-data-visualization-project-using-r-71bc1b9a7495. LastaccessedDecember11,2020.

[2] TheBike-sharingBlog.Availableonlinehttp://bikesharing.blog-spot.com/.Lastaccessed December11,2020.at

[3] X.Zhao,M.Han,Q.Deng,K.Xue.Decisionsonsupplyanddemandofbicycle-sharing servicebasedoncobweb model.

[4] J.Büttner,H.Mlasowsky,andT.Birkholz.OptimisingBikeSharingin European Cities,unpublished,2011.

[5] L.Wang,C.Wang.Analyzingbikerepositioningstrategiesbasedonsim-ulations for Public Bike sharing systems presented at 2013 Second IIAI In-ternationalConferenceon AdvancedAppliedInformatics.

[6] A. Kadri, I. Kacem and K. Labadi. A branch and bound algorithm forsolvingthe staticrebalancingprobleminbicyclesharingsystem.

[7] L. Chen et al. Bike sharing station placement leveraging heterogeneousurban open data, in Proceedings of the 2015 ACM International Joint Con-ference on Pervasive and Ubiquitous Computing UbiComp '15, 2015, pp.571-575

[8] Capitalbikesharingdatasets.Availableonlineathttps://s3.amazo-naws.com/capitalbikesharedata/index.html.accessedDecember12,2020.Last

[9] A.Ramesh,S.Nikhil.Avaibleonline:https://github.com/ArthiRamesh/bike-sharing-demandprediction.LastaccessedDecember12,2020

[10] L. Wang, C. Wang. Analyzing bike repositioning strategies based onsimulations for Public Bike sharing systems presented at 2013 Second IIAIInternationalConferenceon Advanced AppliedInformatics.

[11] H. Xu, J. Ying. An improved GRASP for the bike sharing re-balancingproblem presented at 2017 International Conference on SmartGridandElectricalAutomation.

[12] Y. Duan and J. Wu. Optimizing the crowdsourcing-based bike stationrebalancing scheme presented at 2019 IEEE 39th International ConferenceonDistributedComputing Systems (ICDCS).

[13] Y. Li, Y. Zheng. Citywide bike usage prediction in a bike sharing sys-tem.

[14] S. Feng, H. Chen, C. Du, J. Li and N. Jing. A Hierarchical demand pre-diction method with station clustering for bike sharing systempresentedat2018IEEEThirdInternationalConferenceonDataScienceinCyberspace.

[15] A.F.Imani,R.Hampshire,L.MarlaandN.Eluru.Anempiricalanalysisof bike sharing usage and rebalancing: Evidence from Barcelona andServilleY.Li,Y.Zheng.Citywidebikeusagepredictioninabikesharing system.

[16] D.A.Freedman.StatisticalModels:TheoryandPractice.Cam-bridgeUniversityPress.p.26.Asimpleregressionequationhasontheright

[17] H. Xu, J. Ying. An improved GRASP for the bike sharing rebalancingproblempresentedat2017InternationalConferenceonSmart GridandElec-trical Automation