Predicting Biker Density at Bikeshare Station Intersections in San Francisco

Mahika Dubey
Computational Media
UC Santa Cruz
Santa Cruz, USA
mahika@ucsc.edu

Alan Peral Ortiz
Computer Science
UC Santa Cruz
Santa Cruz, USA
aperalor@ucsc.edu

Rakshit Agrawal
Computer Science
UC Santa Cruz
Santa Cruz, USA
ragrawa1@ucsc.edu

Angus G. Forbes
Computational Media
UC Santa Cruz
Santa Cruz, USA
angus@ucsc.edu

***Abstract*—Bike sharing platforms are becoming increasingly common alternatives to public transportation in cities, improving accessibility to areas not reachable by bus, train, or tram. While this can be beneficial for improving city connectivity, it also increases the likelihood of biker related accidents and vehicle collisions, especially in areas where protected bike lanes and safety infrastructure are not already in place. We compare machine learning models to predict biker density at road intersections in the city of San Francisco, using publicly available trip data from the city's most widely used bikeshare service, Ford GoBike, evaluating our model performance by monitoring mean squared error. Alongside our predictive models we develop a heatmap visualization application to display our predictions, providing an additional mode of interaction for users to access the forecasted information. The intended usage of our work is to predict areas of highest biker density at different times so that drivers and bikers can experience improved shared road safety. The deployment of our models can also inform city planning and alternative public transportation development.**

***Index Terms*—machine learning, prediction methods, neural networks, smart transportation, public transportation, bicycles, user interfaces, visualization**

I. Introduction

Bicycle involved accidents make up a significant percentage of reported road collisions in San Francisco. As of the most recent collisions reported from the San Francisco Municipal Travel Agency (SFMTA) Report, 17 percent of all accidents resulting in injuries involve cyclists, a statistic that has doubled in the last decade despite being one of the least reported type of traffic incidences in the city [1], [2]. A recent report even listed San Francisco as the 4th deadliest city for cyclists in America, demonstrating that while the statistics in San Francisco are reason enough for a call to action, biker fatalities are increasing in frequency in cities all over the country [3]. With the volume of bikers sharing roadways increasing yearly, it is imperative for public safety to find ways to alleviate the risks involved.

With this increase in the volume of bikers, and the spread of the technology-enabled sharing economy, services such as Ford GoBike program have emerged that enable rich data collection around usage of bikes. Availability of such data allows for us to build machine learning based systems that can leverage this information and use it for increasing the safety of bikers on the streets. In this paper, therefore, we propose a machine learning solution for forecasting areas of high bike activity at given times, and aim to contribute to the 11th United Nations Global Goal, Sustainable Cities and Communities, by improving the safety of biking as an alternative option for transportation (Target 11.2) [4].

We present generalizable learning models based on neural networks for forecasting the number of bikers at a given bikeshare station at a certain point in time. We evaluate such models using a standard regression learning metric of mean squared error (MSE), which measures the difference between the predicted and actual number of bikers. These models are developed for individual Ford GoBike stations in the city of San Francisco, with a focus on the most frequently used stations. Training our models on 24 hour long sequences, we are able to predict the next hour of biker density with an average training and testing mean squared error of 0.00501 and 0.00939 respectively. We compare the performance of our simple regression networks with a more advanced model for ordered time-series predictions, using Long Short Term Memory (LSTM) recurrent neural networks, to find a slightly better average mean squared error of 0.00403 for training and 0.00899 for testing. In addition to our trained and tested models, we provide a map-based visualization web application that displays next-hour predictions for activity density at Ford GoBike stations throughout the city, incorporating locations of known collisions involving bicycles.

The rest of this paper details our process and evaluation methods. We summarize related existing literature on bike usage, safety and traffic prediction research in section II. In section III, we describe further our data collection pipeline for Ford GoBike usage and San Francisco bike-related collisions, followed by the machine learning model used for predicting biker density. We provide a detailed evaluation of these models in section IV, along with the visualization tool developed, followed by a discussion in section V. We finally conclude the paper in section VI with a note on the future usability of such methods.

II. Related Works

A. Bicycle Usage Research

Previous research on bicycle safety has often focused on helmet usage, roadway infrastructure, and other psychological, socioeconomic, and environmental factors [5]–[7]. In

978-1-7281-1780-5/19/$31.00 ©2019 IEEE

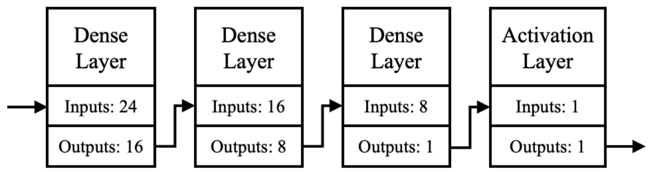

Fig. 1. Regression network architecture for predicting biker density at a given Ford GoBike station.

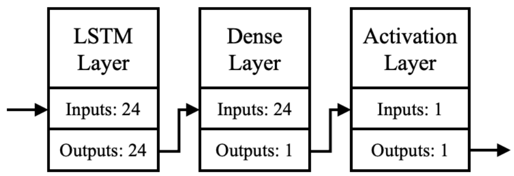

Fig. 2. LSTM network architecture for predicting biker density at a given Ford GoBike station.

a previous survey of bicycle commuting literature, it was identified that predicting and influencing bicycle use could not be accurately described by existing models for evaluating other forms of transportation [6]. Ziari and Khabiri [8] tackled the bicycle safety issue by developing a GIS tool that solely utilized collision reports in Tehran to identify areas of high crash occurrence. However, this tool did not provide an understanding of bicycle usage patterns and thus was limited in its contribution as a planning and development tool.

Responsible planning and infrastructure development have become increasingly important with the rise in popularity of bikeshare programs. The United Nations Department of Economic and Social Affairs' Committee on Sustainable Development described safety issues as one of the challenges in adopting and maintaining bikeshare systems [9]. Some of these difficulties include the fact that bikeshare programs do not typically provide helmets for use and the cyclists that use these services can be inexperienced riders. Given the current state of these deployed services, it is essential to develop tools and software that can influence safer policy and urban development design choices for motorists, cyclists, and pedestrians alike.

B. Traffic Prediction

Traffic flow prediction and analysis has been the focus of intensive research across numerous domains. Min and Wynter [10] detail the use of LSTM neural networks to predict traffic speed using road microwave sensor data in Beijing, China. In Los Angeles County, real world data and spatio-temporal events like rush hour and accidents are used to make traffic prediction more accurate [11]. Research has been conducted to develop a framework to incorporate social media data into existing traffic prediction methods to make longer-term traffic prediction more reliable [12]. There are a variety of additional methods for traffic prediction that seek to make the predictions more precise, immediate, and robust over longer windows of time [10], [13], [14]. However, these approaches, while helpful, are not amenable to direct use by the various stakeholders involved in creating safer roadways for cyclists and pedestrians. Furthermore, the focus on motor vehicle traffic prediction is unsuitable for bicycle usage prediction. This combination of existing issues reveals a gap in our understanding of roadway and cyclist relations that we would like to work towards closing.

III. Methods

A. Data

The SMFTA expanded their bikeshare program as Ford GoBike in 2017, providing comprehensive e-bike access for city residents and visitors, after having received great success in their four year pilot study [15]. Monthly trip data for all Ford GoBikes in the San Francisco Bay Area are made available to the public as CSV files in the format of the North American Bike Share Association's (NABSA) General Bike Feed Specification (GBFS) [16]. We use data for bike stations located within the city of San Francisco, from June 2017 through March 2019 for our experiments. While many residents of San Francisco own the bikes that they use, a high percentage of residents are subscribers of Ford GoBike [15]. Therefore, we assume usage statistics for the bikeshare program to be fairly representative of biking trends in the city.

In order to create a clean data set for building our learning models, we performed a pre-processing round on the raw CSV data. We first performed a filtering to limit the data to San Francisco. We used the latitude and longitude of the stations and removed the ones falling outside the range of San Francisco. Ford GoBike has been deployed outside of San Francisco in San Jose, Oakland, Berkeley, and Emeryville, and therefore it was important to clean the information from additional cities when limiting the learning to San Francisco. The movement patterns of cities vary based on their demographic, industry, and geography. We, therefore, removed all references to stations in these other cities. Once we isolated the San Francisco trip data, we aggregated the number of recorded visits to each station by hour, where a visit is counted if a trip either started or ended at that station within that hour. There were some additional columns in the raw CSV data that we did not use for this study, such as the year of birth of the rider, bike id, and whether or not the customer was a subscriber. We did not consider these features to be important for the task of generating biker density predictions. Many stations did not have recorded visits at certain times, such as between 1 a.m. and 5 a.m. when general bicycle usage was sparse, and some stations were not located in very populated areas leading to a small number of active users. In order to ensure that we did not have any missing entries, we substituted these missing hours with entries that log 0 visits. This process generated a total of 15336 hours of visit data for all 175 Ford GoBike stations in San Francisco. These per-station usage totals were then split into normalized training, validation, and testing sets (10761, 1517, and 3058 sequences respectively for every bikeshare station) and reshaped to generate ordered sequences of 24 hour

TABLE I
MEAN SQUARE ERROR FOR MODELS FOR 25 MOST FREQUENTED FORD GOBIKE STATIONS IN SF

Station Name	Regression Training	Regression Testing	LSTM Training	LSTM Testing
Market St at 10th St	0.00379	0.01091	0.00206	0.01067
Powell St BART Station (Market St at 4th St)	0.00719	0.01509	0.01931	0.04542
Powell St BART Station (Market St at 5th St)	0.00390	0.00964	0.00317	0.00827
Berry St at 4th St	0.00376	0.01084	0.00167	0.00517
Montgomery St BART Station (Market St at 2nd St)	0.00361	0.00946	0.00172	0.00647
Civic Center/UN Plaza BART Station (Market St at McAllister St)	0.00571	0.01271	0.00453	0.01213
The Embarcadero at Sansome St	0.00490	0.00581	0.00356	0.00464
San Francisco Caltrain Station 2 (Townsend St at 4th St)	0.00226	0.00663	0.00110	0.00442
San Francisco Ferry Building (Harry Bridges Plaza)	0.00576	0.00926	0.00279	0.00589
Embarcardero BART Station (Beale St at Market St)	0.00375	0.00708	0.00217	0.00465
Union Square (Powell St at Post St)	0.00728	0.01340	0.00648	0.01296
Steuart St at Market St	0.00454	0.00777	0.00247	0.00488
Townsend St at 7th St	0.00506	0.00903	0.00289	0.00598
3rd St at Townsend St	0.00193	0.00293	0.00141	0.00253
Golden Gate Ave at Polk St	0.00552	0.00575	0.00463	0.00529
San Francisco Caltrain (Townsend St at 4th St)	0.00446	0.00529	0.00198	0.00433
Valencia St at 24th St	0.00717	0.00947	0.00523	0.00821
Post St at Kearny St	0.00651	0.01217	0.00465	0.00994
Broadway at Kearny	0.00618	0.01177	0.00538	0.01049
Valencia St at 16th St	0.00616	0.00813	0.00537	0.00822
The Embarcadero at Bryant St	0.00671	0.00996	0.00551	0.00992
Beale St at Harrison St	0.00377	0.00935	0.00236	0.00937
Howard St at Beale St	0.00578	0.00949	0.00232	0.00458
17th St at Valencia St	0.00407	0.00876	0.00364	0.00849
Folsom St at 3rd St	0.00545	0.01402	0.00425	0.01178
Average	***0.00501***	***0.00939***	***0.00403***	***0.00899***
Median	***0.00506***	***0.00946***	***0.00317***	***0.00821***

usage totals and the actual number of visits to each station in the next hour.

We also utilize traffic collision data as reported by the California Highway Patrol (CHP) through the Statewide Integrated Traffic Records System (SWITRS) [17]. The SWITRS data is available upon request, and is provided in CSV format containing a variety of information including collision date, nearest intersection, and whether the collision involved a bicycle. The data contains 76 columns of related information, but for our purposes we only selected a small number of these: the collision date, time, primary and secondary roads (intersection), the collision severity, the number of people killed, the number of people injured, the latitude, and the longitude. Additionally, in correspondence to the objective of the paper, we only looked at collisions that involved cyclists. We believe that these were essential data points to include in our visualization tool, as discussed in Section IV-B.

B. Neural Networks

Neural networks are mathematical models built to mimic the structure and learning process of neural connections in the brain. These models can be trained to learn highly complex, nonlinear relationships in data, and were a natural fit for the predictive regression problem we aim to solve [18], [19]. We built two different types of models for each bike station to learn to make predictions for the next hour of biker density based on the recorded usage statistics of the previous 24 hours. This not only gives us accurate predictions based on recent trends in usage, but also provides models that can evolve quickly with long term changes as data is collected and re-training occurs. The first model we use is a multi-layered network consisting of 3 fully connected dense layers with a tapering number of units and a final activation layer on the output as seen in Figure 1. We then constructed an additional model based on the sequential learning approach of neural networks using a Long Short Term Memory (LSTM) network [20]. LSTMs are suited for time series prediction tasks as they are designed specifically to learn from ordered sequences. The structure of our LSTM based model can be seen in Fig. 2.

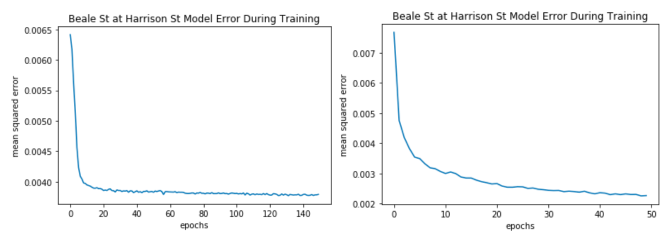

Fig. 3. Decreasing mean squared error (MSE) for Regression model training (left), and LSTM Network training (right) for the Beale St at Harrison St Ford GoBike station.

We built both the networks in Keras [21], a python library for neural networks and deep learning, and ran the training and testing experiments on GPU-accelerated Google Colab notebooks. Both models used the mean squared error as loss functions and Adam [22] for optimization. Training time for

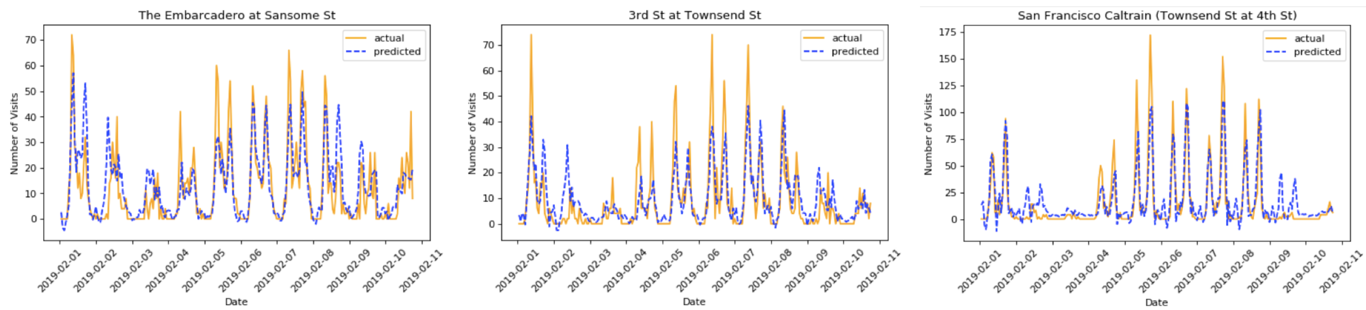

Fig. 4. Comparison of the actual number of biker visits to 3 different stations over a subset of our test data (the first 10 days of February 2019) to the number of bikers predicted by the regression model.

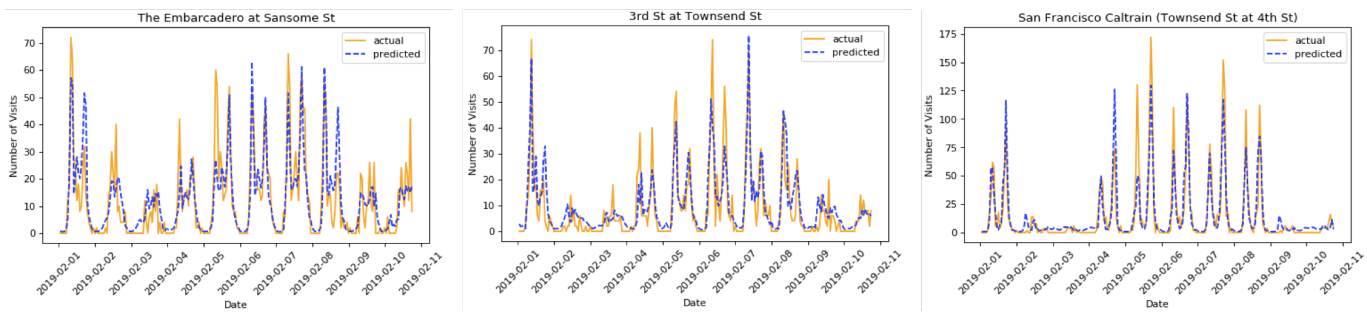

Fig. 5. Comparison of the actual number of biker visits to 3 different stations over a subset of our test data (the first 10 days of February 2019) to the number of bikers predicted by the LSTM model.

each station was 15 minutes for the regressive model over 150 epochs and about 1 hour over 50 epochs for the LSTM. We discuss the performance of the networks in detail in the next section.

IV. Evaluation

A. Model Accuracy

The learning problem for our model deals with a time-series regression. In order to judge the accuracy for such objectives, a commonly used metric is the mean squared error. In terms of performance, a model with a lower mean squared error is a better model. In Figure 3, we see the mean squared error over the course of training for a single station for both the regression model and the LSTM based model. While the mean squared error of the regression network is initially lower than that of the LSTM, the LSTM network generally converges to a smaller error in fewer epochs than needed by the regression network. This suggests that a larger number of epochs for the LSTM may potentially result in even better accuracy for predicting usage.

Once the models were trained, we evaluated their error in predicting bike density for unseen sets of data. The testing errors of both models for each of the 25 most frequented Ford GoBike stations in SF are summarized in Table I. On average, with our regression models we see a mean squared error of 0.00501 for training and 0.00939 on testing data. In most cases, the LSTM outperforms the associated regression model, with average mean squared errors of 0.00403 and 0.00899 for training and testing respectively. The improvement in accuracy in the LSTM over the regression model is even more apparent when comparing the results for individual stations.

Figure 4 and figure 5 plot the actual and predicted number of bikers at certain stations over a 10 day period for both model types. As can be observed from the plots, the regression models are great at capturing times of high and low biker density, including the different usage patterns over weekends and weekdays, reflecting the broad range of training sequences. In consistence with the results on mean squared error, the LSTM models outperform the regression models in all cases, having learned to predict the large variances in usage over time to a much more accurate level. We can return to the difference in the number of epochs used for training each of the models in analyzing their performance. Given that the LSTM was trained over only 50 epochs compared to the 150 used for the regression models, the plots visually support the intuition that the LSTM approach could benefit further from continued training over a larger number of epochs. Results from both models prove that accurate biker density prediction is possible and quite simple to implement. Given the potential impact to reduce casualties by providing road-users with such information we hope our approach can be integrated into existing systems to better inform communities about such patterns.

B. Heatmap Visualization

We developed a visualization tool to accompany our predictive models in this study. We utilized the Ford GoBike data to create a heatmap layer using the Leaflet JavaScript library

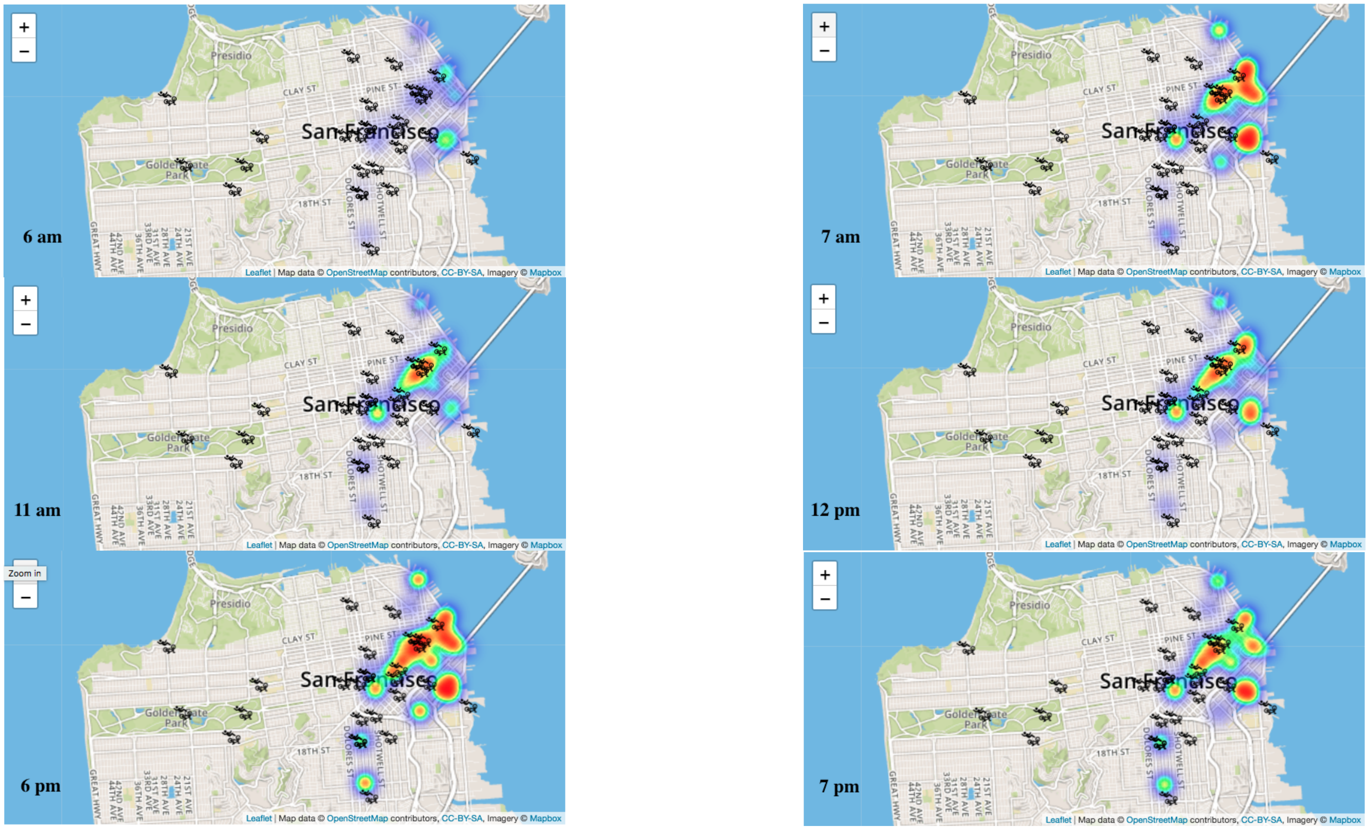

Fig. 6. Bicycle usage predictions generated by our regression models for 25 stations in San Francisco. We see that our model predicts higher usage during the morning and afternoons, mirroring real time rush hour patterns.

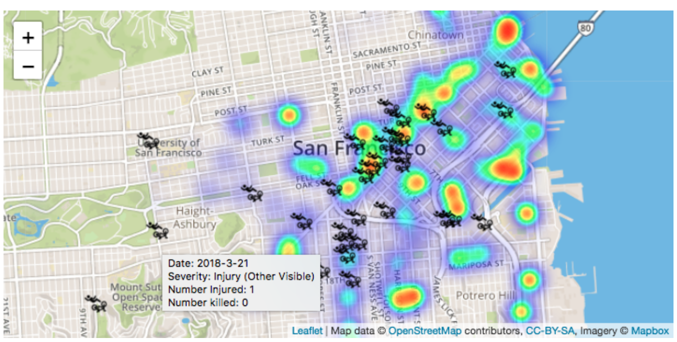

Fig. 7. A snapshot of our visualization. The heatmap indicates the number of visits for a given station, with higher frequencies being indicated by warmer colors, while individual accidents are represented by the cyclist tool-tips. Hovering over the fallen cyclists reveals additional information about the accident.

designed for building interactive maps and the MapBox API for the map design. We allow the user to select any date and time (to the nearest hour) within the ranges for which we have the data. In rounding the bike docking and renting times to the nearest hour, general patterns in usage can be observed, which would have been harder to notice if we specified the minutes as well. In this heatmap, higher frequencies are represented by warmer colors (red, yellow) while lower frequencies are represented by cooler colors (blue, green).

In addition to the heatmap layer, we also added tool-tips that show where bicycle accidents occurred. Due to the scarcity of the bicycle collision data (there were 1302 data points from June 2017 to March 2019), we aggregate the data by month. Additionally, as most of the collision data provided by the California Highway Patrol (CHP) only included intersection information and not coordinates, we had to geocode the data using the Geocoder Python library and the MapQuest API. This allowed us to then plot the accidents on the map as well. Hovering over the fallen cyclist icons on the tool reveals additional information that details the date of the accident, the severity of the accident, the number of people injured, and the number of people deceased (as reported by the CHP). This design allows users to see points of high bicycle usage frequency and relate it to collision data, identifying particularly dangerous locations for cyclists in the city. Figure 7 shows our visualization in use, demonstrating the heatmap and the accident tool-tips. Users decide the date and time they wish to see data for by using a simple HTML range-input control scheme.

This tool also allows us to visualize our predictions in understandable ways. Figure 6 show our generated predictions for 25 stations at six distinct hours on the same day. We can see that the small amount of predicted bicycle usage at 6 a.m.

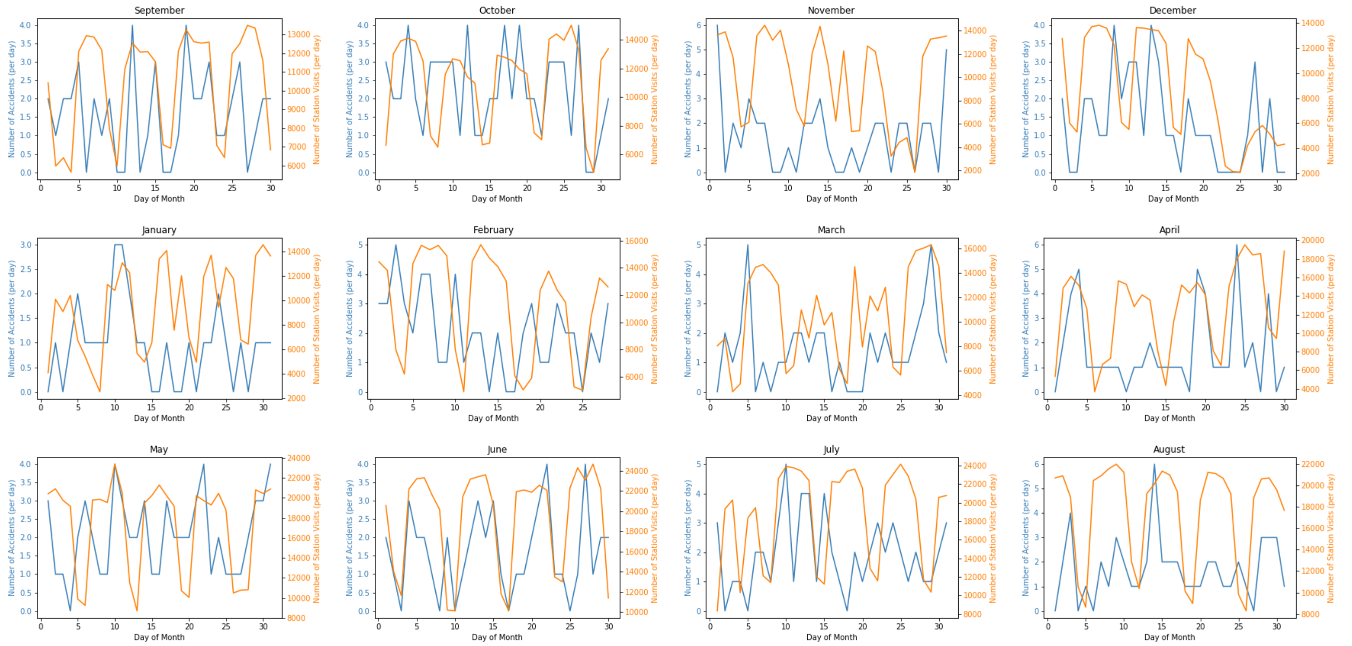

Fig. 8. Daily GoBike bicycle usage and accident data in San Francisco from September 2017 to August 2018.

starts to increase significantly at 7 a.m., drops slightly between 11 a.m. and 12 p.m., and increases once more in the late afternoon before tapering off again. This pattern mirrors rush hour traffic, suggesting that it may be especially important to ensure cyclists' safety during those hours.

V. Discussion

In completing this project we sought to develop suitable ways for our prediction data and insights to be usable by people and organizations. As alternative forms of safe and easily accessible public transportation are encouraged across cities around the world, we look to develop frameworks for thinking about shared road safety.

The lack of data-driven approaches with regards to bicycle safety, coupled with personal dangerous cycling interactions with motorists motivated us to examine the relationship between bicycle usage and accident occurrence. Figure 8 shows the total daily number of visits for every GoBike station in San Francisco as well as the daily number of accidents from September 2017 to August 2018. We can see that the increases and decreases in bicycle usage are very closely mirrored by the rise and fall in accident occurrences. This trend is even more alarming when we consider the problem of under-reporting in bicycle related accidents in San Francisco [2]. Increased accident counts could bolster the observed relationship between biker density and accident occurrence in the city, further encouraging the use of tools like these for the purposes of informed decision making. Providing easily interpretable data through our visualization tool to the relevant stakeholders can lead to the increased protection of vulnerable road users.

Our visualization web application provides a way for users to view predicted biker traffic in San Francisco and reported locations of bike-involved collisions, and can be used as a standalone application for monitoring such trends. Our predictions could also be served through an API for anyone to access and use within their own applications, for instance, with a mapping or routing application. We use latitude and longitude information for each Ford GoBike station to plot traffic over the map. This approach can easily be applied to other cities as long as collisions information and bikeshare usage trends are available in similar formats, making our model and approach completely generic. Further, even in the absence of bikeshare usage, data available from bike tracking services, and health monitoring devices that track the ride information can also be used in the same way with our methods.

Popular navigation applications such as Google Maps and Garmin Satellite GPS systems provide real-time vehicle traffic data for roads, including providing warnings about individual events such as slowdowns caused by accidents, road closures due to construction work, or dangerous conditions including roadside fires or flooding. These features give drivers the ability to choose alternate routes to avoid traffic and be prepared for unusual situations. Especially in big cities, sharing the roads with bicyclists leads to congestion and reduced safety. If navigation applications could warn drivers of intersections on their journey with forecasts of high biker density, and provide alternate routes to avoid them, it is possible that frequency of collisions or even just excessive traffic jams could be mitigated, particularly during busy times. With our proposed API for serving prediction data, navigation applications can simply integrate this information and provide the information within their apps.

VI. Conclusion and Future Work

In this paper, we presented a biker density prediction model and a visualization tool to help improve biker safety. Our current results for the city of San Francisco show true potential in the space of predictive modeling for bikeshare trends and station usage, work that directly contributes to the UN Global Goal 11.2 of improving sustainable cities through the development of better systems of public transportation [4]. Given the standardized format of available data (NABSA GBFS), our solution can be generalized to cities all over the US to help combat the rising issue of cyclist fatalities [3]. It is important that such data be incorporated into predictive analytic systems to shape infrastructure policy, building preventative solutions to the improve shared road safety. The usage of our forecasting models can serve as a first step towards reducing the frequency and severity of bicycle accidents in cities overall, promoting a safer and more accessible alternative form of public transportation. In developing machine learning solutions for predicting bike traffic, we explored different architectures for robust and accurate prediction models, but further experiments with additional hyperparameter tuning including the addition of more complex layers, different optimization algorithms, and more epochs for training can likely lead to improved accuracy and longer prediction sequences.

There is great value to be derived in simplifying the creation of visualizations like the one presented in this paper. While bikeshare data may have a standardized format, collision data may vary wildly based on the reporting entity. This lack of collision data reporting standards presents both a technical and social challenge that must be overcome for systems such as ours to benefit users in an impactful way. Active and responsible reporting of incidents involving bicycles on roadways can support the development of dependable applications for viewing and monitoring areas with a greater likelihood of unsafe conditions, making our roads safer for all.

References

[1] B. Loo and K. Tsui, "Bicycle crash casualties in a highly motorized city," *Accident; analysis and prevention*, vol. 42, pp. 1902–7, 11 2010.

[2] S. Sustainable Streets Division. [Online]. Available: https://www.sfmta.com/sites/default/files/reports/2016/San Francisco Collisions Report 2012 2015.pdf

[3] A. Wang, "Cyclist deaths are rising nationwide - here are the 20 most dangerous cities to ride a bike in," Jul 2019. [Online]. Available: https://www.businessinsider.com/most-dangerous-cities-to-ride-a-bike-in-2019-7

[4] "Goal 11: Sustainable cities and communities." [Online]. Available: https://www.globalgoals.org/11-sustainable-cities-and-communities

[5] B. H.-Y. Lee, J. L. Schofer, and F. S. Koppelman, "Bicycle safety helmet legislation and bicycle-related non-fatal injuries in california," *Accident Analysis Prevention*, vol. 37, no. 1, pp. 93 – 102, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001457504000600

[6] E. Heinen, B. van Wee, and K. Maat, "Commuting by bicycle: An overview of the literature," *Transport Reviews*, vol. 30, no. 1, pp. 59–96, 2010. [Online]. Available: https://doi.org/10.1080/01441640903187001

[7] T. A. Petritsch, B. W. Landis, H. F. Huang, and S. Challa, "Sidepath safety model: Bicycle sidepath design factors affecting crash rates," *Transportation Research Record*, vol. 1982, no. 1, pp. 194–201, 2006. [Online]. Available: https://doi.org/10.1177/0361198106198200124

[8] Z. H and M. M. Khabiri, "Applied gis software for improving pedestrian bicycle safety," *Transport*, vol. 20, 08 2005.

[9] P. Midgley, "Bicycle-sharing schemes: enhancing sustainable mobility in urban areas," *United Nations, Department of Economic and Social Affairs*, vol. 8, pp. 1–12, 2011.

[10] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, pp. 606–616, 08 2011.

[11] B. Pan, U. Demiryurek, and C. Shahabi, "Utilizing real-world transportation data for accurate traffic prediction," in *2012 IEEE 12th International Conference on Data Mining*, Dec 2012, pp. 595–604.

[12] J. He, W. Shen, P. Divakaruni, L. Wynter, and R. Lawrence, in *IJCAI International Joint Conference on Artificial Intelligence*, 2013, pp. 1387–1393.

[13] J. Park, D. Li, Y. L. Murphey, J. Kristinsson, R. McGee, M. Kuang, and T. Phillips, "Real time vehicle speed prediction using a neural network traffic model," in *The 2011 International Joint Conference on Neural Networks*, July 2011, pp. 2991–2996.

[14] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, April 2015.

[15] K. Smith, "Bikeshare," May 2019. [Online]. Available: https://www.sfmta.com/getting-around/bike/bike-share

[16] I. Motivate International, "System data." [Online]. Available: https://www.fordgobike.com/system-data

[17] C. H. Patrol, "Statewide integrated traffic records system." [Online]. Available: http://iswitrs.chp.ca.gov/Reports/jsp/userLogin.jsp

[18] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, 2016.

[19] K. P. Murphy, *Machine learning: a probabilistic perspective*. The MIT Press, 2012.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980