Enhancing Bike-sharing Demand Forecasting: Anomaly Detection and Feature Selection in LSTM Networks

Pedro Nunes
School of Design, Management and Production Technologies Northern Aveiro
University of Aveiro Oliveira de Azeméis, Portugal
e-mail: pnunes@ua.pt
José Paulo Santos
Department of Mechanical Engineering,
University of Aveiro Aveiro, Portugal
e-mail: jps@ua.pt

***Abstract*—Accurate forecasting of casual bike-sharing demand is crucial for optimizing operations and resource allocation. This study employs a Long Short-Term Memory (LSTM) network to predict hourly bike rentals, incorporating temporal, meteorological, and categorical features. To enhance the model, we integrate an anomaly detection step using the Local Outlier Factor (LOF) method, treating its output as an additional feature. The initial LSTM model achieved a Root Mean Squared Error (RMSE) of 34.26. Incorporating anomaly detection based on weather-related data, such as temperature and humidity, and subsequently removing those features, led to an improved RMSE of 30.86. Feature permutation analysis was then used to assess variable importance. The most critical predictors were whether the day was a working day and which working day it was, highlighting clear behavioral patterns in casual bike-sharing demand. By combining anomaly detection with feature selection, we enhance the interpretability of LSTM-based forecasting models, which are often considered black boxes. Removing redundant features simplifies the model while potentially improving accuracy, making it more transparent and efficient. These findings provide valuable insights for bike-sharing system operators, enabling data-driven decision-making for demand management and operational planning.**

Keywords-LSTM; Bike-Sharing; Feature permutation; Anomaly detection; Interpretability.

I. INTRODUCTION

Accurate forecasting of the demand for bike sharing is essential to optimize operations and improve urban mobility [1]. Various factors influence bike-sharing demand, including built environment characteristics, weather conditions, and temporal trends [2][3]. For example, the authors in [2] demonstrated the impact of urban infrastructure and land use on ridership levels, while [4] investigated the operational challenges associated with bike redistribution to balance demand across stations. Additionally, event detection techniques have been utilized to identify anomalies in bike-sharing data, enhancing forecasting accuracy by accounting for unexpected fluctuations in usage [5].

Recent advancements in Machine Learning (ML) have enabled the development of sophisticated predictive models, such as Deep Learning (DL) approaches, to capture the complex temporal and spatial dependencies inherent in bike-sharing usage patterns [3]. Among these, Long Short-Term Memory (LSTM) networks have shown promise in time-series forecasting due to their ability to model long-term dependencies in sequential data [1]. The authors in [6] conducted a comparative study between multiple linear regression and LSTM models, finding that LSTM significantly outperformed traditional regression techniques in predicting bike-sharing demand when considering time and weather factors.

Recent studies have demonstrated the effectiveness of ML techniques in capturing the complex, non-linear relationships inherent in bike-sharing data. For instance, [7] employed an artificial immune system combined with an Artificial Neural Network (ANN), to predict bike-sharing demand. Similarly, [8] proposed a Spatial-Temporal Graph Attentional LSTM approach that integrates multi-source data, including historical trip records and weather information, to enhance short-term demand predictions. On the other hand, [9] emphasized the importance of analyzing and visualizing bike-sharing demand with outliers, proposing methodologies to model baseline temporal usage patterns and detect significant deviations.

In this study, we employ an LSTM-based approach to predict hourly bike rentals, incorporating temporal, meteorological, and categorical features. To enhance model performance, we integrate an anomaly detection step using the Local Outlier Factor (LOF) method, treating its output as an additional feature. This approach aligns with previous research that highlights the importance of addressing demand variability through advanced modeling techniques [4]. Furthermore, we implement a feature permutation analysis to assess the importance of variables in order to understand the most influential parameters on bike-sharing demand.

One of the key features of the proposed solution is the improvement of the interpretability of LSTM-based forecasting models, which are often seen as black boxes. In realistic contexts, it is of utmost importance to have transparent predictions and to understand the main parameters under the predicted values to optimize operations and decision-making.

The remainder of this document is organized as follows: Section II presents the problem being addressed and describes the dataset used. Section III outlines the proposed combination of anomaly detection and an LSTM network for forecasting bike-sharing demand. Section IV presents and discusses the obtained results. Finally, section V concludes the work.

II. Dataset and problem description

The dataset used in this study originates from the Capital Bikeshare system in Washington, D.C., covering a two-year period from 2011 to 2012. This dataset, originally compiled by Fanaee-T and Gama [5], includes rental data aggregated hourly, and integrates multiple sources of information, including weather data and calendar-based attributes, to provide a comprehensive view of bike rental patterns. The dataset attributes include:

- Temporal attributes: Date (dteday), season (season), year (yr), month (mnth), hour (hr for the hourly dataset), weekday (weekday), and working day (workingday).
- Weather conditions: Weather situation (weathersit), temperature (temp), apparent temperature (atemp), humidity (hum), and wind speed (windspeed).
- Rental information: Count of casual users (casual), registered users (registered), and the total count of rented bikes (cnt).

The primary objective of this study is to develop an accurate model to forecast the hourly bike rental demand for casual users, using an LSTM-based approach. Since this demand is influenced by several factors, such as weather conditions, holidays, and special events, it is intended to incorporate a method for feature importance, in order to assess the factors that most contribute to the model's prediction, thus mitigating the black box nature of most DL models.

III. Methodology

The proposed methodology, depicted in Figure 1, encompasses a preprocessing step that structures the raw data for training the LSTM model. After training, a feature importance analysis is conducted to identify the most relevant features. This process yields a predictive model that forecasts hourly bike-sharing demand while providing insights into the most significant contributing factors.

A. Preprocessing

The preprocessing stage encompasses several steps, as depicted in Figure 2. The dataset includes both numerical and categorical variables. To prepare the data for training the LSTM model, one-hot encoding was applied to transform all categorical variables into numerical representations. Note that the numerical variables are already normalized, and for this reason, there was no need to scale them.

Anomalies and event-driven variations, such as unusual spikes or drops in bike rentals, may arise due to special events or extreme weather conditions [1] [2]. Considering this, the LOF algorithm [10] was employed to identify anomalies in the weather-related data (temperature, humidity, perceived temperature, and windspeed). LOF is an unsupervised anomaly detection method that identifies outliers based on local density variations relative to their neighbors. It quantifies how isolated a data point is by comparing its density to that of surrounding points. The variables used for anomaly detection included:

- *temp:* Normalized temperature;
- *atemp:* Normalized apparent temperature;
- *hum:* Normalized humidity;
- *windspeed:* Normalized wind speed.

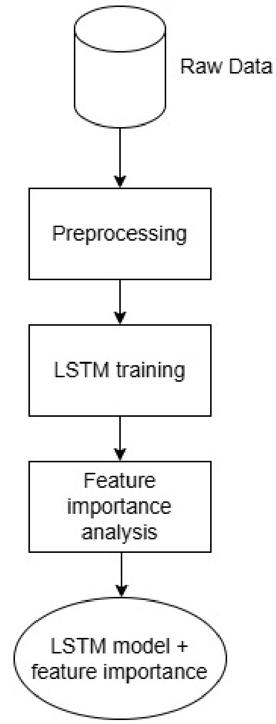

Figure 1. Overview of the proposed methodology.

In this study, a neighborhood size of 24 points was selected, corresponding to the time window used in the LSTM model, as will be further discussed. The method was implemented using the Scikit-learn library, with a contamination ratio of 0.05.

The LSTMs are a type of Recurrent Neural Networks (RNNs) designed to capture long-range dependencies in sequential data, making it well-suited for time-series forecasting tasks. Unlike traditional RNNs, which suffer from vanishing gradient problems when learning long-term dependencies, LSTMs incorporate specialized gating mechanisms to regulate the flow of information [11].

In this study, the input to the model consists of time-ordered sequences of features extracted from the dataset, including weather conditions, temporal attributes. Each training instance is structured as a rolling window of 24 consecutive hourly observations, where the model uses data from the previous 24 hours to forecast the bike demand for the next hour.

After structured, the dataset was divided in training, validation, and testing instances, maintaining the temporal order, as illustrated in Figure 3. The proportion used was 68-12-20 for training, validation, and test, respectively. By maintaining the temporal order we want to ensure that no data leakage occurs during the training process.

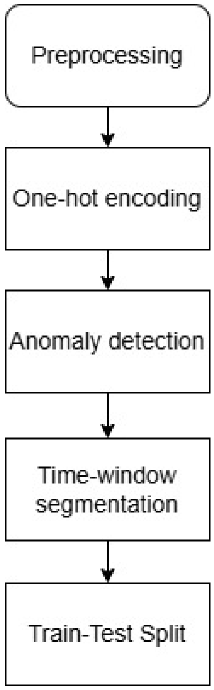

Figure 2. Overview of the proposed methodology.

Figure 3. Overview of the data split, maintaining the temporal order. 68% training data, 12% validation data, and 20% test data.

B. LSTM model training

The time-window size, N_W, for the LSTM was set has 24, since it is reasonable to assume time dependencies of the last day, to forecast casual users of bike-sharing. Figure 4 depicts the structure of the proposed LSTM model. It uses temporal and weather features from the current hour and from the last 24 hours. The output of anomaly detection is also used as feature, but in this scenario, the weather-related features used to compute it, were removed. The LSTM network has one layer with 25 cells, and it is followed by three dense layers with 50,20, and 1 neurons, respectively. After the,e first LSTM, and dense layers, a dropout of 0.1 was used. This model was implemented using Keras and TensorFlow, Python libraries.

We use the Adam optimizer and mini-batch of 32 samples, to optimize the weights and bias of the DL model. The adopted learning rate was 0.001 if the number of epochs was lower than 10, and then decreased according to $l_r(i) = l_r(i-1) * e^{-0.01}$, where $l_r(i)$ is the learning rate of the current epoch. This strategy was chosen to stabilize the training process, leading to better fine-tuning of the model. The maximum number of epochs was set to 100. Note that, to avoid overfitting, the early stop is applied after 10 consecutive epochs with no improvement in the validation score.

C. Feature importance analysis

The feature importance analysis quantifies the contribution of each feature to the LSTM model's predictive performance. We employ the permutation importance technique, which assesses feature relevance by randomly shuffling the values of a given feature and measuring the resulting decline in model performance. The greater the degradation, the more critical the feature is to the model.

One advantage of this method is its model-agnostic nature, meaning it can be applied to any trained estimator. Additionally, by performing multiple permutations, we obtain a measure of variance in the importance scores, enhancing result reliability.

Feature importance is computed by (1), where i_j is the importance of feature j, s is the reference score for the model (e.g., F1-score for classification or RMSE for regression), and K denotes the number of permutations. In this work, we set $K = 5$.

$$i_j = \frac{1}{K} \sum_{k=1}^{K} s_{k,j} - s. \tag{1}$$

IV. Results and discussion

After obtaining the first LSTM model, by using all features except the anomaly detection output, we evaluated its performance on the test dataset using Root Mean Square Error (RMSE) as the primary metric. Figure 5 displays the predicted and actual values for the number of casual bike-sharing users over the first 10 days of the test period. To enhance visualization, only 240 hours of the test data are shown. The RMSE for the full test set is 34.25, which is reasonable given the range of values observed for casual user counts.

To assess the impact of incorporating anomaly detection, we removed the weather-related features originally used to train the LOF model and instead included the anomaly detection output as an input feature. The model was then retrained. As shown in Figure 5, this revised approach improved the model's performance, reducing the RMSE to 30.86. Additionally, it lowered computational complexity by using fewer input features.

Accurately predicting bike-sharing demand is crucial for optimizing urban mobility decisions. However, beyond predictive accuracy, understanding which factors most influence predictions is essential for informed decision-making. To achieve this, we applied the conditional feature permutation method to evaluate the importance of each input variable. First the correlation matrix was computed for the features, and then features with a correlation higher than 0.75 were shuffled conditionally, to assure that dependencies between features are not broken during the process.

As depicted in Figure 6, the most influential features in the trained model are $workingday$ (indicating whether a day is a weekday or weekend) and the features $weekday\_n$, which identify the day of the week. The seasons of the year

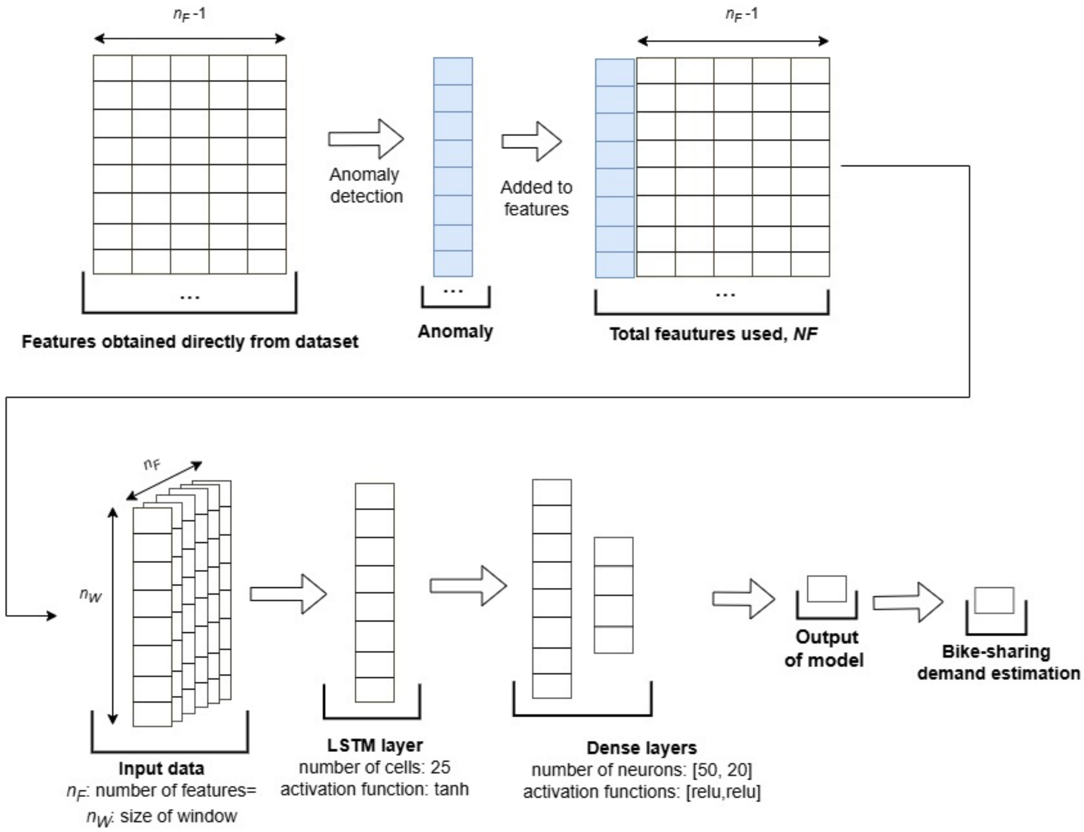

Figure 4. Proposed LSTM architecture.

and general weather conditions $weathersit$ appear to be less significant.

It is interesting to note the features such as temperature (tem), perceived temperature ($atemp$), and wind speed ($windspeed$) have very low importance, since anomaly detection also has low importance, however the anomalies detected using these features as basis have a reduce the computational complexity of the model, while improving its performance. This suggests that instead of absolute weather values, what matters most is whether the weather conditions at a given hour deviate significantly from recent patterns.

Following the initial evaluation, we removed four features (temperature, perceived temperature, humidity, and windspeed), and replaced them by the anomaly detection output. The refined model achieved an improved performance, with an RMSE of 30.86 (compared to 34.26), as depicted in Figure IV. This indicates that combining DL with feature importance analysis and anomaly detection allows us to:

- Identify the most influential features driving the predictions.
- Reduce model complexity by eliminating less relevant variables.
- Maintain comparable predictive performance while using fewer features.

V. CONCLUSION AND FUTURE WORK

This study proposed an LSTM-based approach to forecast hourly bike-sharing demand, incorporating anomaly detection and feature importance analysis. Integrating the LOF method allowed the model to account for unexpected variations in demand, while the feature permutation analysis enabled the identification of the most influential predictors. Results demonstrated that the most critical features were related to the day of the week and whether it was a working day, confirming clear behavioral patterns in casual bike-sharing usage.

Furthermore, the feature selection step reduced model complexity while improving predictive accuracy. This highlights the potential of combining deep learning with explainability techniques to enhance both performance and interpretability in time-series forecasting tasks.

Future work could explore advanced interpretability methods for deep learning models, such as SHAP (Shapley Additive Explanations) and Integrated Gradients, to provide deeper insights into feature contributions. Additionally, investigating attention mechanisms in LSTM or Transformer-based models could improve both transparency and predictive accuracy. Expanding the methodology to different bike-sharing systems and urban contexts would also help validate its applicability and robustness.

Another promising line of research involves the integration

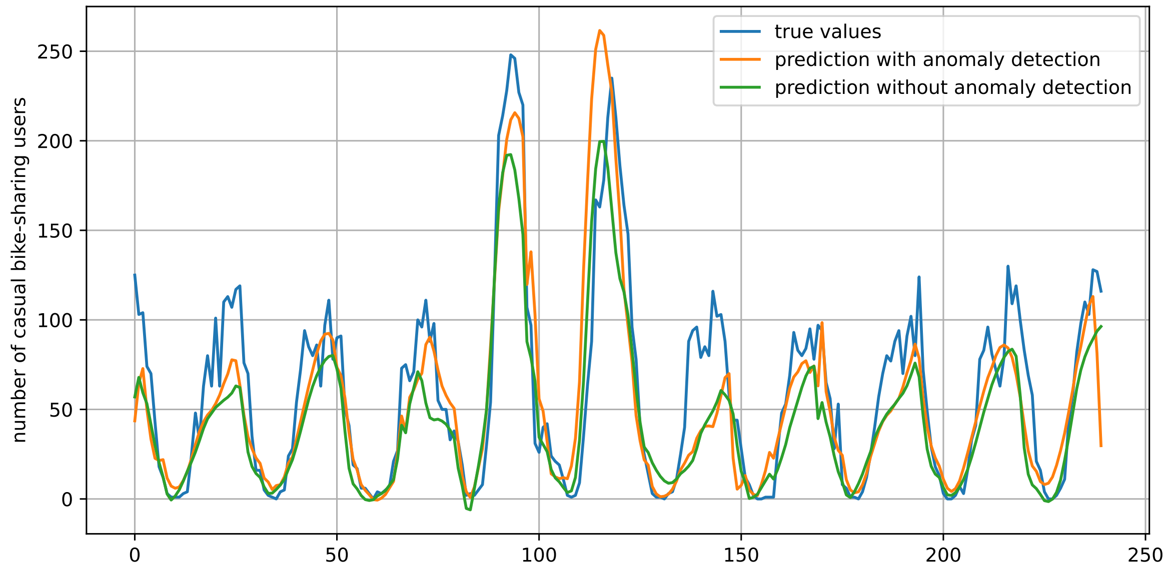

Figure 5. Comparison between prediction scenarios with anomaly detection, and with no anomaly detection for the first 10 days of test data (240 hours).

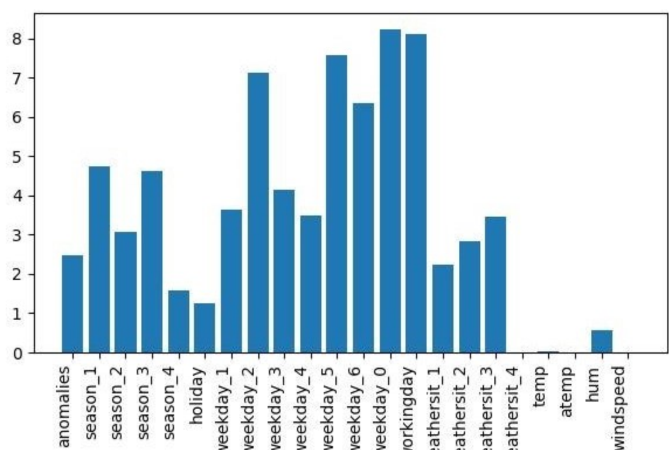

Figure 6. Feature importance obtained through conditional feature permutation.

of IoT modules directly into bike-sharing systems, enabling the real-time collection of weather-related data, such as temperature and humidity, as well as automated user counting, as proposed by [12]. Combined with additional sensors like accelerometers and GPS, this setup could offer valuable insights into user preferences and mobility patterns. Such data could support the development of real-time, context-aware route recommendation systems for cyclists, as explored in [13].

ACKNOWLEDGMENTS

This study was developed within the scope of Project AM2R [C644866475-00000012 – Project No. 15], funded by the PRR – Recovery and Resilience Plan under the Next Generation EU from the European Union. Laboratory support was provided by the projects UIDB/00481/2020 and UIDP/00481/2020, financed by FCT – Fundação para a Ciência e a Tecnologia (DOI: 10.54499/UIDB/00481/2020 and DOI: 10.54499/UIDP/00481/2020).

REFERENCES

[1] “Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach,” *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, Dec. 2018, ISSN: 0968090X. DOI: 10.1016/j.trc.2018.10.011.

[2] W. El-Assi, M. S. Mahmoud, and K. N. Habib, “Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in toronto,” *Transportation*, vol. 44, pp. 589–613, 3 May 2017, ISSN: 15729435. DOI: 10.1007/s11116-015-9669-z.

[3] E. Eren and V. E. Uz, “A review on bike-sharing: The factors affecting bike-sharing demand,” *Sustainable Cities and Society*, vol. 54, p. 101 882, Mar. 2020, ISSN: 22106707. DOI: 10.1016/j.scs.2019.101882.

[4] T. Raviv, M. Tzur, and I. A. Forma, “Static repositioning in a bike-sharing system: Models and solution approaches,” *EURO Journal on Transportation and Logistics*, vol. 2, pp. 187–229, 3 Aug. 2013, ISSN: 21924376. DOI: 10.1007/s13676-012-0017-6.

[5] H. Fanaee-T and J. Gama, “Event labeling combining ensemble detectors and background knowledge,” *Progress in Artificial Intelligence*, vol. 2, pp. 113–127, 2-3 Jun. 2014, ISSN: 2192-6352. DOI: 10.1007/s13748-013-0040-3.

[6] X. Zhang, Z. Yu, and C. Xie, "Optimizing bike sharing demand prediction: A comparative study of multiple linear regression and lstm models based on time and weather factors," *Journal of Education, Humanities and Social Sciences*, vol. 42, pp. 353–359, Dec. 2024, ISSN: 2771-2907. DOI: 10.54097/c2934z45.

[7] P.-C. Chang, J.-L. Wu, Y. Xu, M. Zhang, and X.-Y. Lu, "Bike sharing demand prediction using artificial immune system and artificial neural network," *Soft Computing*, vol. 23, pp. 613–626, 2 Jan. 2019, ISSN: 1432-7643. DOI: 10.1007/s00500-017-2909-8.

[8] X. Ma, Y. Yin, Y. Jin, M. He, and M. Zhu, "Short-term prediction of bike-sharing demand using multi-source data: A spatial-temporal graph attentional lstm approach," *Applied Sciences (Switzerland)*, vol. 12, p. 1161, 3 Feb. 2022, ISSN: 20763417. DOI: 10.3390/app12031161.

[9] N. Rennie, C. Cleophas, A. M. Sykulski, and F. Dost, "Analysing and visualising bike-sharing demand with outliers," *Discover Data*, vol. 1, p. 1, 1 Mar. 2023, ISSN: 2731-6955. DOI: 10.1007/s44248-023-00001-z.

[10] M. M. Breuniq, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," *SIGMOD Record (ACM Special Interest Group on Management of Data)*, vol. 29, no. 2, pp. 93–104, 2000. DOI: 10.1145/335191.335388.

[11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 08997667. DOI: 10.1162/NECO.1997.9.8.1735.

[12] P. Nunes, C. Nicolau, J. Santos, and A. Completo, "From a traditional bicycle to a mobile sensor in the cities," SCITEPRESS - Science and Technology Publications, Jul. 2020, pp. 81–88, ISBN: 978-989-758-419-0. DOI: 10.5220/0009349700810088.

[13] P. Nunes, A. Moura, and J. Santos, "Solving the multi-objective bike routing problem by meta-heuristic algorithms," *International Transactions in Operational Research*, vol. 30, pp. 717–741, 2 Mar. 2023, ISSN: 0969-6016. DOI: 10.1111/itor.13114.

 ISBN: 978-1-68558-286-9