Machine Learning Methods for the Volumetric Analysis of Unicamp's University Restaurants

Fausto Marques Pinheiro Junior* **Giovana Melo dos Santos**† **Guilherme Philippi**‡
Lucca da Paz§ **Samara Viriato Vilar Dias**¶
Dept of Applied Mathematics, IMECC, UNICAMP,
13083-859, Campinas, SP
E-mail: f249213@dac.unicamp.br, g240268@dac.unicamp.br, g299081@dac.unicamp.br,
l290774@dac.unicamp.br, s297354@dac.unicamp.br

ABSTRACT

The uncertainty of daily demand in university restaurants is a challenge for ensuring adequate resource allocation and minimizing waste. To address this, we developed an integrated dataset and applied Machine Learning models to predict user volumetry in Unicamp's university restaurants. Our work involved data collection, database construction, and the implementation of three supervised learning algorithms (Neural Network, Support Vector Machine and Random Forest) to support the process of planning for the decision maker.

After the data collection phase, we constructed a database that contains not only information about the restaurants itself but is also accompanied by other categories: temporal data and weather data to aid in the desired prediction of the Machine Learning algorithms applied to the problem. Alongside the database, we designed the package `Rice.jl`[1] implemented in the `julia` programming language to support, manage and enhance the training of Machine Learning algorithms over the dataset.

Because our exploratory analysis revealed strong temporal dependencies among the variables, LSTM was selected as one of the predictive models, given its ability to retain information from previous observations through memory. The LSTM model was implemented in Python using the `PyTorch` library due to its robustness and flexibility for deep learning applications.

Support Vector Machines (SVM) constitute a family of algorithms widely used for classification and regression. In essence, SVMs determine a separating hyperplane that maximizes the margin between data classes. In our study, SVM regression was implemented using the `LIBSVM.jl` package in conjunction with `MLJ.jl`.

Decision Trees are supervised learning models that apply a branching structure to classify or predict outcomes. Their nodes contain decision rules based on variables values, while leaf nodes correspond to class labels or continuous predictions. For this work, we used a random forest regressor, an ensemble of multiple decision trees, implemented via the `DecisionTree.jl` package alongside `MLJ.jl`.

We present in Figure 1 an illustrative comparison between the real and predicted series for the lunch demand at the university restaurant RU restaurant by the LSTM algorithm. Although some local variations are observed, it captures the overall consumption dynamics, confirming its ability to model patterns in the university restaurant system.

---

*PhD student CNPq
†PhD student FAPESP (Grant No. 2024/21644-8)
‡PhD student FAPESP (Grant No. 2025/19235-5)
§master's student CAPES (Grant Finance Code 001)
¶master's student CNPq (Grant No. 404102/2022-0)
[1]`https://gitlab.com/gphilippi/rice/`

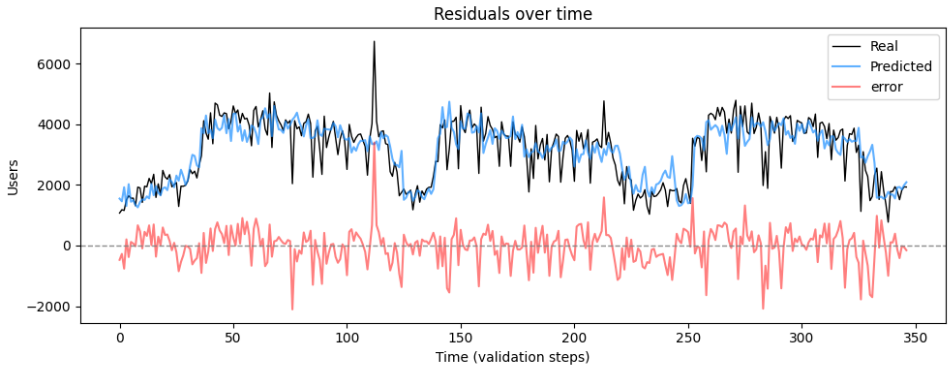

Figure 1: Example of real vs. predicted user demand for RU lunch using the LSTM model

Table 1: Model Performance

	RMSE			**MAE**		
Model	**RU**	**RA**	**RS**	**RU**	**RA**	**RS**
SVM						
Lunch	0.671	0.546	0.371	0.525	0.389	0.291
Dinner	0.513	0.412	0.232	0.409	0.295	0.19
RF						
Lunch	0.770	0.815	0.766	0.61	0.634	0.603
Dinner	0.770	0.815	0.766	0.61	0.634	0.603
LSTM						
Lunch	0.63	0.72	0.53	0.44	0.63	0.49
Dinner	0.62	0.79	0.55	0.49	0.66	0.45

To support the comparative analysis, were ranked performance metrics using a color scale ranging from red to blue ( ), where darker blue values indicate better performance and red worse performance. This visualization in Table 1 highlights the relative strengths of each model across multiple restaurants and each meal.

The experiments revealed, as expected, differences among the models. We observed that LSTM architecture achieved the best performance in capturing temporal correlations, showing strong interpretation of consumption patterns. The Random Forest model demonstrated robustness and interpretability but lacked sensitivity to temporal ordering. Meanwhile, SVM, although capable of modeling nonlinear relationships, showed limitations in handling the high-dimensional categorical features and long-term dependencies inherent to the dataset.

Keywords: *Machine Learning, Time Series, Prediction*

References

[1] Breiman, Leo, Random Forests, *Machine Learning Springer Science and Business Media LLC, Philadelphia*, 45 (2001) 5-32

[2] Cortes, Corinna and Vapnik, Vladimir, Support-vector networks, *Machine Learning Springer Science and Business Media LLC*, 20 (1995) 273-297.

[3] Hochreiter, Sepp and Schmidhuber, Jürgen,Long Short-Term Memory, *Neural Computation MIT Press*, 9 (1997) 1735-1780.