



Machine Learning Applied to Walk-in Demand Prediction of a University Counseling Center

Erin Magee, Meserret Karaca, Michelle Alvarado, Ernesto Escoto, and Alvin Lawrence

University of Florida

Faculty mentor: Michelle Alvarado, Industrial & Systems Engineering

Abstract

University of Florida Counseling and Wellness Center (UF CWC) employs a walk-in appointment policy for emergency counselling needs. The attendance to UF CWC has increased in walk-in appointment traffic every year since data collection began in 2010, averaging a 7% increase in patient visits per year. However, demand for walk-in services is highly uncertain on an hourly, daily or weekly basis. Additionally, emergency needs of students should be met immediately before they become catastrophic. Thus, demand prediction becomes an important aspect to dynamically schedule counselors to deal with unexpected demand scenarios. This project provides data visualization and utilizes machine learning techniques to predict future demand to assist with scheduling. We identified seasonal trends in historical visit data from the center, including peaks at the beginning of semesters and around finals. We then used the visit data to train a Gradient Boosting algorithm to predict demand. This model predicted demand with a mean of 4.2 patients per hour and mean square error of 1.75. Our results contributed to better demand prediction models for the UF CWC in order to better support student needs with adequate staffing levels.

Keywords: counseling, machine learning, demand prediction, appointments

Introduction

In Florida, the budgets allocated to mental health resources are unable to match the increasing demand and complexity of student problems being seen at university counseling centers (Dunkelberger, 2017). This increasing demand is coupled with more complex mental health issues in students, like changing psychosocial conditions in today's college populations, i.e. higher rates of depression diagnoses in millennials (Watkins, Hunt, & Eisenberg, 2011). In addition, counseling centers at universities are already constrained for resources and may lack authority to bring more into the center due to budgets coming down from university administration. In order to address the challenge of balancing resource utilization and meeting demand, some counseling centers are considering a switch to a fully on-call appointment

scheduling model (Shaffer et al., 2016), such as the University of Florida's Counseling and Wellness Center (UF CWC).

In the traditional counseling center scheduling model, clients arrive in-person and complete both an intake form and an intake appointment with a counselor. The traditional model includes scheduled appointment times in order to efficiently utilize counselor time. In some cases, scheduled appointments run longer than expected or patients don't show-up for appointments, resulting in increased wait times or resource underutilization.

In contrast, the on-call model allows clients to come in when they experience severe symptoms (e.g. suicidal thoughts, grief counseling) where they are immediately seen by a triage counselor to assess their needs. Following triage services, clients are given a follow-up appointment or proceed to intake counseling. On-call availability at a university counseling center allows students to see counselors when their symptoms are at their most acute, bringing their needs to the counselor's attention when they are the most severe. Additionally, on-call triage may lessen clinical exacerbation from long wait times and increase client satisfaction (Shaffer et al., 2016).

Previous work has addressed modelling techniques for uncertainty in demand for on-call appointment types in emergency department settings (Bard and Purnomo, 2005; Becker et al., 2018; Hong et al., 2018; Sherali et al., 2002; and Wiler et al., 2011). Bard and Purnomo (2005) used integer programming methods to analyze the fluctuations in supply and demand for nurses and schedule them accordingly. Becker et al. (2018) used set-covering based integer programming to schedule emergency medical services. They used employee preferences and previous shift schedules to create a schedule to maximize the minimum number of free weekends for employees. Hong et al. (2018) used machine learning to predict patient admission to the hospital after emergency department triage. This team used factors from triage data and patient history data to predict admissions, in turn helping the admitting facility to schedule resources appropriately. This approach used in emergency departments is analogous to UF CWC's situation because patients can come in at any time they experience symptoms. This approach also mimicked the data given to us by the center. However, no previous work has addressed on-call/emergency appointment types in counseling settings. In our paper, we first identified seasonal trends in historical walk-in visit data from the center, including peaks at the beginning of semesters and around finals. We then used the visit data to train a Gradient Boosting

algorithm from machine learning methods to predict demand of walk-in counseling appointments.

University Counseling Center Background

UF CWC currently uses a hybrid on-call model of established appointment times with the option of an on-call counselor available immediately for student emergencies. UF CWC is a premiere center, tied for the third place ranking among university counseling centers nationally (U.S. News, 2019). UF CWC has 41 counselors and sees 30 on-call patients on an average day, with an average of 4.3 patients per hour requesting on-call services. In 2018, the center saw 5,340 clients in total. Roughly 45% of them are on-call appointments, which are the focus on this paper.

We used a policy brief from previous appointment data analysis done at UF CWC and the raw data from the center to look for trends in demand. The counseling center hypothesized that trends show up yearly: as semesters begin, putting new stresses on students, as semesters end, during final exam weeks, as well as those times within the semester when students return from breaks. Additionally, the counseling center observed external factors which could affect the demand for services such as football games, severe natural events like hurricanes, and traumatic on-campus events, such as school shootings or on-campus suicides. The center recognizes and acts accordingly with semester trends data, staffing more counselors at the beginnings of semesters and during finals, however unpredictable factors such as a football game outcomes can add even more uncertainty to demand. This project sought to standardize the periods of time where demand becomes the most unpredictable. The time-periods found in this paper and identified through interviews will become the basis for the next part of this research – consistent reactive scheduling around those times. Figure 1 shows the weekly demand for on-call appointments at UF CWC by week of the semester. Each line depicts an academic year's demand, starting in mid-August each year. The patterns for demand were quite stable between academic years. However, the demand consistently increased year to year. It was observed that in 2014-2015 the highest demand was just above 200 appointments in one week and in 2015-2016 the demand frequently exceeded 200 appointments per week. In the most recent year, on-call appointments reached 450 in one week. The variance was very high on a weekly basis due to the academic calendar's breaks and exams.

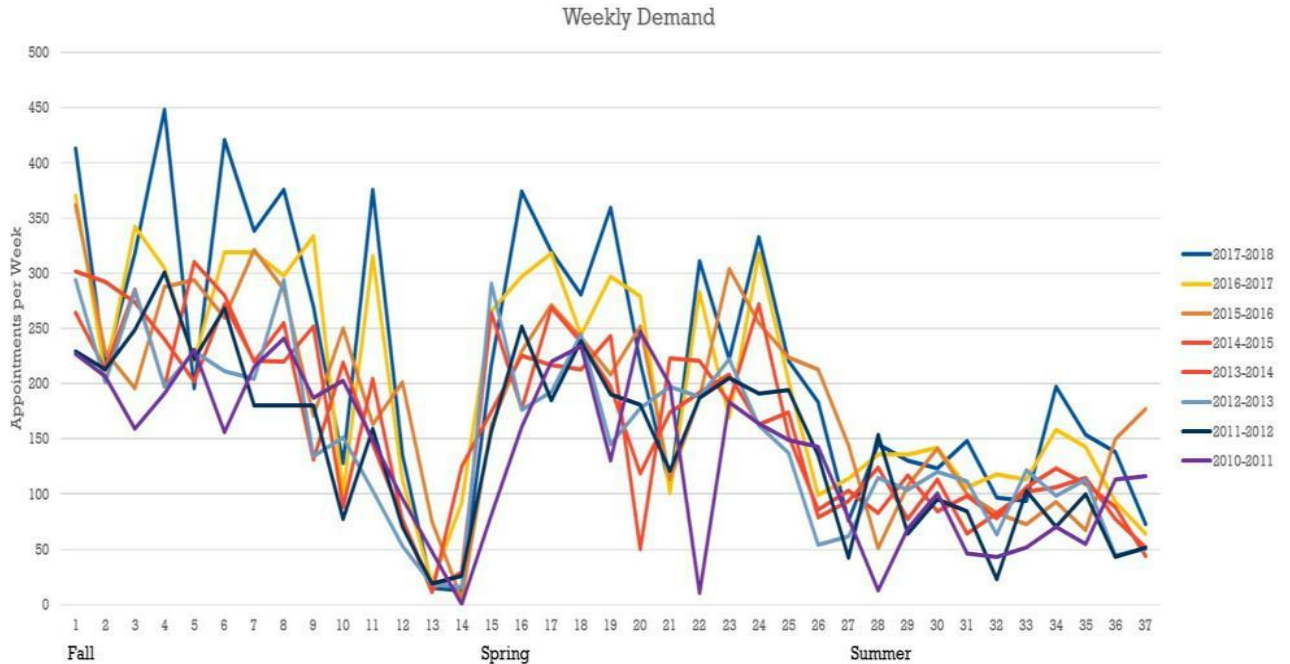


Figure 1 All weekly appointment data at UF CWC from 2010-2018

There were large increases in appointment demand in the beginning of the Fall semester in late August. Drop offs in demand occur shortly after that first week but peak again right before the ends of the semesters (early December for the Fall semester and mid April in the Spring). In addition to these visible trends in demand, the center has been experiencing an overall increased demand for services. This study will assist the center in staffing to meet overall increased demand as well as finding areas for improvement in the center's on-call scheduling procedure, which is currently done manually before the start of each semester.

Method

The objective of this study is to utilize machine learning techniques to find patterns in the data and use the patterns to form inferences about future demand at the center. The goals of this research project are 1.) to understand the current state of walk-in demand and counselor utilization at UF CWC, 2.) to use machine learning to identify patterns in demand, and 3.) to build a tool to predict demand at UF CWC to better staff counselors to meet student emergency needs.

As a first step, we used Excel to visualize the given data and verify certain patterns in demand. We found the patterns the counselors had alluded to, specifically the steep hike in demand in the

first week of school and the subsequent peaks as students faced events such as exams and football losses. Once we identified peaks and set a baseline for the general shape we expected, we explored machine learning options for finding an optimal prediction model. In general, supervised learning takes in a set of training data, learns the function to map inputs to outputs, finds patterns and inferences in data, and uses those patterns to predict outcomes. Data sets are typically split between training and testing data (e.g. 80% training data and 20% testing data). The training data is used to fit parameters of the model. This will include input values and output values, also called target values. The testing data is used to provide an evaluation of the model fit from the training dataset. The team chose a gradient boosting machine learning algorithm to compare on- and off-campus events with demand at the counseling center.

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, a differentiable loss function $L(y, F(x))$, number of iterations M .

Algorithm:

1. Initialize model with a constant value:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma).$$

2. For $m = 1$ to M :

1. Compute so-called *pseudo-residuals*:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \dots, n.$$

2. Fit a base learner (e.g. tree) $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $\{(x_i, r_{im})\}_{i=1}^n$.

3. Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)).$$

4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. Output $F_M(x)$.

Figure 2 Gradient Boosting Algorithm (Bouquin, 2018)

Gradient boosting utilizes regression trees with updating starting points to find optimal patterns in data. Each time a regression tree is made, the loss function, here defined as the mean square error, is recalculated and a new starting point for a subsequent regression tree is found. Pseudo-code for the gradient boosting algorithm is provided in Figure 2. The model updates

itself when it finds a more optimal starting point, resulting in better predictions as it iterates (Gorman, 2017). We chose gradient boosting because of this ability to update the model and make better predictions from previous models. As seen in Figure 3, the gradient boosting algorithm is like hitting a golfball toward the hole over many iterations with better starting points. In this analogy, the player can update the way the ball is hit to get it closer to the hole with each try. Similarly, the gradient boosting algorithm updates its starting points to get closer to its goal of smallest mean square error. In Figure 3 below, y is the value or function the algorithm is trying to approximate, F_0 and subsequent F represent the estimates, and each delta represents the difference of the differences of the true value and estimate. The algorithm is seeking the least difference between the estimates and the true value, seen graphed on the curve of mean square error loss function.

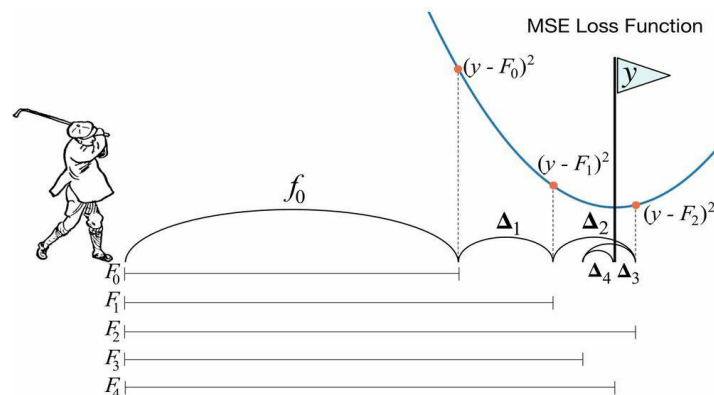


Figure 3 Illustration of Gradient Boosting with mean square error (MSE) loss function (Parr, n.d.)

We incorporated factors given to us by the counseling center in our model: football game wins/losses and whether they were home or away games, major natural events, and final exam dates. These factors were highlighted by counseling center administration in interviews as factors with the most influence on the uncertainty in demand. In addition to these factors obtained through counseling center interviews, we added in factors such as time of year (first week of semester, second week, etc.), day of the week, and hour of the day. All of these factors acted as weights to the regression tree updating predictions.

Table 1 List of Factors included in gradient boosting algorithm.

Factors That May Influence Demand	
1. Weather	2. Semester of the Year
3. Football Games	4. Week of the Year
5. Holidays/Semester Breaks	6. Day of the Week
7. Exams	8. Hour of the Day
9. Time of the Year	10. Demand in Previous Days

We trained the model on the first seven years of data, from academic years 2010-2011 until 2016-2017, and used the remaining academic year, 2017-2018, as a test set for the model.

Results

Results showed that our prediction model could mimic the trends seen in previous years. As seen in Figure 4, our model predicted the same trends throughout a year, though it is lacking in the scale of appointments it predicts. This model predicts more conservatively the number of appointments occurring throughout the academic year. The mean square error associated with this model was 1.75 appointments. Our model created good though conservative predictions for demand. We saw the biggest discrepancy at high peaks in demand, such as Sept. 2017 and Feb. 2018, related to a hurricane and school shooting event that directly affected the state of Florida and UF students.



Figure 4 Real vs. Predicted appointment data for academic year 2017-2018

Figure 4 shows the comparison of our model to real data for the entire academic year 2017-2018. Below, Figure 5 shows a week with very close predictions. This is typically a week following the Spring Break, so we think there is minimal variability due to a lack of academic, football, or weather related factors.

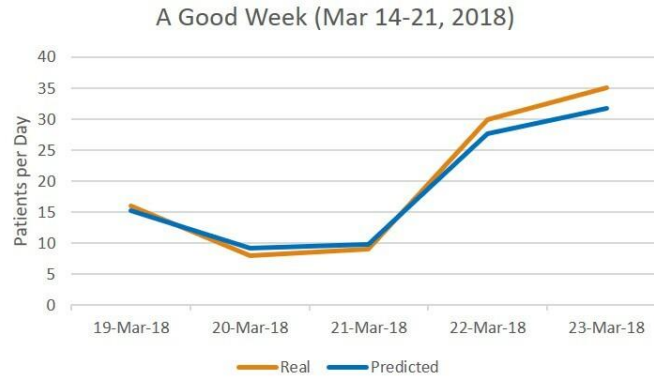


Figure 5 Results for the week of March 14-21, 2018

Figure 6 shows data for the week leading up to Homecoming (October 6-7th, 2017). This was one of the weaker predictions by the model because it consistently predicted lower than observed demand. We think this is due to the increased variability from football and academic related factors at that time.

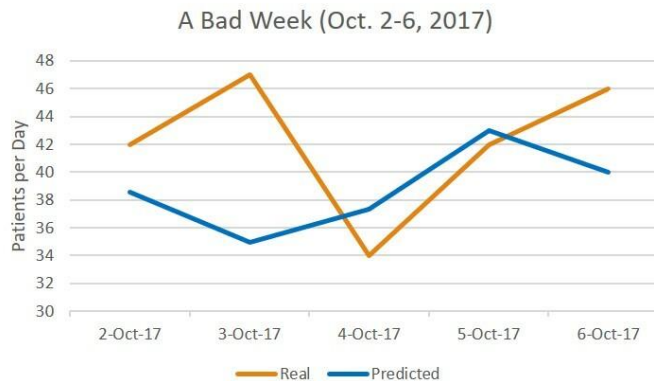


Figure 6 Results for the week of October 2-6, 2017

Figures 7 and 8 below illustrate the model's ability to predict on an hourly basis. In Figure 7, we've shown one of the better predicted days of the past academic year. The model matched the steep rise leading up to lunch and the sudden plateau seen around lunchtime. However, the model did not match as well when the demand decreased in the afternoon.

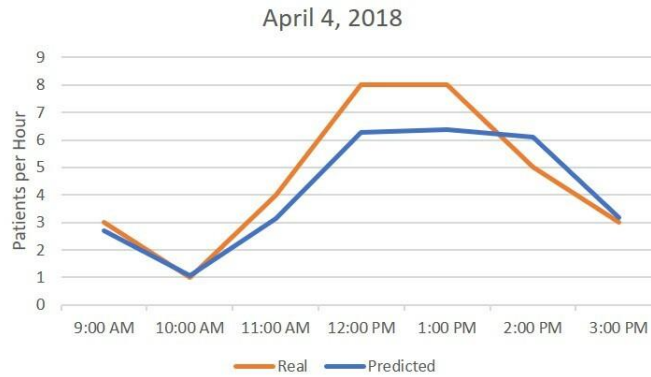


Figure 7 Results from April 4, 2018

Figure 8 shows one of the worst prediction days. The model under-predicted for the morning, making too conservative predictions for both 9:00 am and 10:00 am hours. It then over-predicted in the afternoon based on the higher demand in the morning, predicting around 5 patients per hour when we actually saw 4 and under.



Figure 8 Results from September 11, 2017

Conclusion

In conclusion, this work found the seasonal trends in the appointment data the counselors had alerted us about. Data cleaning and visualization showed us the trends the counselors had pointed out and we were able to standardize those trends in the model prediction results. The gradient boosting model was able to predict hourly patient demand of 4.2 patients per hour with a mean square error of 1.75. When the factors that affect the uncertainty in demand were manipulated, we found the combination with the least error included all of the factors. We found that days of

the week and week of the semester had the most positive impact on our predictions. Other machine learning models should be tested as well. In future work, we will be developing an interactive tool that incorporates predictions and previous data with counselor schedules and resource utilization calculations. We presented this information to the UF CWC administrators and are also looking for ways to include counselor preferences and vacation into the scheduling tool.

Acknowledgements

The authors would like to thank the University of Florida Counseling and Wellness Center for proposing this research and providing data.

References

- Bard, J. F., & Purnomo, H. W. (2005). Short-Term Nurse Scheduling in Response to Daily Fluctuations in Supply and Demand. *Health Care Management Science*, 8(4), 315-324. doi:10.1007/s10729-005-4141-9
- Becker, T., Steenweg, P. M., & Werners, B. (2018). Cyclic shift scheduling with on-call duties for emergency medical services. *Health Care Management Science*. doi:10.1007/s10729-018-9451-9
- Bouquin, D. (2018, December 7). Stochastic Gradient Boosting Machines: The basics. Retrieved March 22, 2019, from https://slides.com/dbouquin/mljc_2018#/0/21
- Dunkelberger, L. (2017, March 27). Universities Push For More Mental Health Funding. Retrieved from <https://www.wuft.org/news/2017/03/27/universities-push-for-more-mental-health-funding/>
- Gorman, B. (2017, January 24). A Kaggle Master Explains Gradient Boosting. Retrieved from <http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/>
- Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *Plos One*, 13(7). doi:10.1371/journal.pone.0201016
- Parr, T., & Howard, J. (n.d.). How to explain gradient boosting. Retrieved from <https://explained.ai/gradient-boosting/L2-loss.html>
- Shaffer, K. S., Love, M. M., Chapman, K. M., Horn, A. J., Haak, P. P., & Shen, C. Y. (2016). Walk-In Triage Systems in University Counseling Centers. *Journal of College Student Psychotherapy*, 31(1), 71-89. doi:10.1080/87568225.2016.1254005
- Sherali, H. D., Ramahi, M. H., & Saifee, Q. J. (2002). Hospital resident scheduling problem. *Production Planning & Control*, 13(2), 220-233. doi:10.1080/09537280110069667
- U.S. News, (2019). *The Best Education Schools for Student Counseling and Personnel Services, Ranked*. U.S. News. Retrieved from <https://www.usnews.com/best-graduate-schools/top-education-schools/student-counseling-rankings>
- Watkins, D. C., Hunt, J. B., & Eisenberg, D. (2011). Increased demand for mental health services on college campuses: Perspectives from administrators. *Qualitative Social Work: Research and Practice*, 11(3), 319-337. doi:10.1177/1473325011401468

Wiler, J. L., Griffey, R. T. & Olsen, T. (2011), Review of Modeling Approaches for Emergency Department Patient Flow and Crowding Research. *Academic Emergency Medicine*, 18: 1371-1379. doi:10.1111/j.1553-2712.2011.01135.x