

Identifying High-Risk Workspaces during COVID-19 using Machine Learning

Lex Drennan and Jorge Lozano

Steelcase
Grand Rapids, MI
ldrennan, jlozano1@steelcase.com

Matthew Chesser

University of Michigan
Ann Arbor, MI
chesserm@umich.edu

Erin Carrier

Grand Valley State University
Allendale, MI
carrier@gvsu.edu

Abstract

The COVID-19 pandemic has wreaked havoc worldwide, on both public health and the worldwide economy. While necessary, quarantine and social distancing requirements have left many companies unable to reopen their offices in a safe manner. We present a model capable of identifying workspaces at high risk for COVID-19 disease transmission and illustrate how existing techniques for quantifying uncertainty in machine learning can be applied to assess the reliability of these predictions. This model is developed using a dataset created by leveraging historical sales data and detailed product information, and it is in the process of being utilized to identify customers to whom to reach out to facilitate the retrofitting of workspaces to support a safe return to the office.

Introduction

The COVID-19 pandemic has resulted in a wide range of challenges for our society. Beyond the serious, worldwide health impacts, one of the most pressing issues has been the effect of the pandemic and the resulting quarantine on the corporate economy. The inability for employees to return to the office has led to disruptions, inefficiencies, and communication challenges, increasing the financial impact on the world economy. While working from home is one option, it is not for everyone. Not all companies can manage it and not all employees benefit from it.

There is substantial research that shows the power of the office space to unlock creativity, productivity, collaboration, and help mold an organization's culture, e.g. (Nielsen 2016; Hodari 2015). Over the last five years, it is estimated that North American corporations have spent tens of billions of dollars on their office spaces. These spaces, however, were not designed to face the challenges of a pandemic.

To recover the advantages of working together in an office, organizations will need to retrofit their spaces in a way that allows them to be able to bring their employees back into the office in a safe and efficient manner. To understand the current state of the North American office at scale, we utilize new data sources such as digital drawings of office spaces, etc and predict which workspaces are most at risk of COVID transmission. While some corporations may have the ability to analyze their existing workspaces for dis-

ease transmission, this is a difficult and potentially cumbersome task that requires substantial effort, and many organizations may not do so or will need assistance retrofitting their workspaces. Designers of office space and furniture will need to support the identification and retrofitting of workspaces at high risk for disease transmission, and they are well-poised to do so with their extensive knowledge of available products that can be used to adapt existing offices. While these designers are well-equipped to assist with the retrofitting of workspaces, many office furniture companies have incomplete knowledge of the exact setup by a specific client. In this paper, we demonstrate how a machine learning model can be used to detect *high-risk workspaces*, those that are at high risk for disease transmission, from corporate sales records of office furniture products and product information. This model is being deployed to identify clients likely to have high-risk workplaces and help prioritize who is reached out to first to enable organizations to welcome their employees back to a safe work environment.

Although applications of machine learning to COVID related risk detection are prevalent, e.g. (Pal et al. 2020; Wang et al. 2020), the authors are unaware of any application of machine learning to identify risk of furniture layout in office spaces or any other space application. Thus, we believe this to be both an extremely novel and crucially relevant COVID related machine learning application.

Data

To understand if a workspace is set-up in a way that presents a risk to the worker, we need to understand more than just product composition. The risk comes from particular details of how the workstation is set-up in relation to others. For this reason, it is critical to understand key metrics such as distance between workers, height of division elements (such as panels or screens), and relative orientation. To create the dataset for supervised learning, we extract key information from digital office floor plan designs, to process and classify thousands of predefined workstations into high risk and low risk based on their product composition, distance, division, and orientation. This results in a dataset that has very detailed product information and what this means in terms of having a set-up that is considered a high risk for a typical office worker, in terms of disease transmission.

The testing dataset is derived from product-level ship-

Application Type	Risk Classification
120 degree ("120") Other Meeting Space Private Space Work Space	Low Risk
Face-to-face (F2F) In-line/Opposing (ILO)	High Risk

Table 1: Application types and their risk classifications.

ment data containing client information, sale data, delivery location, and installation location within the client’s office. Since the shipment data is at a product-level granularity, business logic and the specificity of location data was used to group products into workstations. A unique identifier for each workstation is constructed to allow the testing dataset to be joined back with the client information after classification. The specific product lines chosen for the training and testing set are determined via business logic.

Each training and testing instance corresponds to a unique workstation, and contains one feature per product line, where each feature corresponds to the percent of the total workstation sale amount attributed to that specific product line. This data is sparse, since no single workstation contained half of the product lines selected.

Model

The ultimate purpose of the model is to predict whether a furniture application is high risk or low risk. To do this, the different application types are identified and a risk class determined for each, which can be seen in table 1.

Due to the imbalance of the training data, with only 9.26% of the training observations being high risk, stratified sampling is necessary. Based on the generally superior performance of random forests in initial model exploration, further exploration is limited to random forests. With stratified sampling, an ensemble model of individually high performing models is trialed. Stage 1 of this model directly predicts risk class. Stage 2 predicts application type from which risk class is derived. In Stage 3 an ensemble vote biased towards high risk is undertaken. The voting rules are:

- If models agree – take model vote
- If models disagree – take high risk vote

This leads to 3 options for high risk and 1 for low risk, creating a 3:1 bias in voting.

Hyperparameter Optimization

To ensure high-risk workspaces are detected, we seek to create a biased model that prioritizes the sensitivity score of “high risk” application types instead of overall model accuracy. This is explored through the optimization of class weights, which penalize the model’s misclassification of “high risk” application types more heavily than misclassification of other application types.

Our final multi-stage ensemble model consists of several constituent random forest models. The individual random forest models are divided into two types of classification:

the classification of the seven application types and the binary classification of the derived “high risk” and “low risk” classes. For all constituent random forest models, the process is as follows:

1. The entire training data set is divided into a stratified 80/20 train/test split.
2. A grid search is performed over each class-weight combination. An integer weight range of [1,10] is used, except for non-high-risk application types which are fixed at 1.
3. For each class-weight combination, 5-fold cross validation is performed and the class of interest’s mean cross-validation sensitivity score is stored. The class of interest for the derived risk class classifiers is the “high risk” class. For the application type classifiers, the classes of interest are ILO, F2F, and a combination of ILO and F2F (for which the sensitivity score is simply the sum of the individual sensitivity scores).
4. The ten models with the highest mean cross-validation sensitivity scores are trained on the entire training split and compared using the testing split. For derived risk class classifiers, the model with the highest high-risk sensitivity score and low-risk sensitivity score above 90% is chosen. For application type classifiers, the model with the highest sensitivity score on the class of interest is chosen.
5. Two additional models are trained with class weights corresponding to inverse proportion of class frequency in the training split and tested using the testing split. The class weight values are calculated using Scikit-learn’s built-in “balanced” and “balanced_subsample” options for “class_weight” hyperparameter.
6. The performances of the models from steps 4 and 5 are compared, and the class weights corresponding to the model whose class of interest sensitivity score is highest on the testing split are chosen as the optimal class weights.

The choice to use an 80/20 train/test split for additional verification is made to prevent overfitting due to the relatively small training dataset ($n \approx 21,000$). The final weights selected are:

- Optimize High Risk: “low risk” : “high risk” relative weighting of 1:6
- Optimize F2F and ILO (Balanced): {"120":1, F2F: 3, ILO: 6, other: 1, meeting: 1, private: 1, workspace: 1}
- Optimize F2F: {"120":1, F2F: 7, ILO: 2, other: 1, meeting: 1, private: 1, workspace: 1}
- Optimize ILO: {"120":1, F2F: 3, ILO: 7, other: 1, meeting: 1, private: 1, workspace: 1}

For each set of optimal weights, a forest of forests is generated to identify which attributes of the training data are most important for predicting results with the specified weightings, producing an ordered list of attributes for each optimization approach.

To identify the best depth, 10-fold cross-validated random forest model is produced for depths 1:25 for each subset of

Weight Set	Optimal Attributes	Best Depth
Optimize High Risk	55	11
Optimize F2F and ILO (Balanced)	20	24
Optimize F2F	34	23
Optimize ILO	77	24

Table 2: Optimal attributes and depth for each possible weight set (may be different across weight sets).

attributes in priority order, i.e. for $n = 1 \dots 82$ priority attributes. The best high-risk sensitivity result from this process is selected to use as the number of priority attributes and random forest depth for the final model, detailed in table 2.

To identify the overall best performing model, a variety of standalone and ensemble approaches are considered with 100 iterations of each model undertaken with different train/test split samples to generate different random forests and results averaged to produce final sensitivities. The high-risk and low-risk sensitivity scores for a selected subset of the considered models are shown in table 3.

The ensemble of Stage 1 (Risk Class) and Stage 2 Balanced & In-line/Opposing produces the overall best results and is therefore selected as the final model. The overall architecture of the model is illustrated in fig. 1. The model incorporates voting at two stages, both with 3:1 bias:

- Stage 2: assigning final stage 2 risk class using inputs from stage 1B and stage 1C
- Stage 3: assigning final risk class using inputs from stage 1A and stage 2

Option (all weighted)	Low Risk	High Risk
Stage 1	92.04%	83.29%
Stage 1 & Ensemble Balanced, F2F & ILO 3-way ensemble vote for stage 2 2-way ensemble vote for stages 1 and 2	94.59%	91.31%
Stage 1 & Ensemble F2F/ILO 2-way ensemble vote for stage 2 2-way ensemble vote for stages 1 and 2	93.65%	93.21%
Stage 1 & Ensemble Balanced & F2F	93.73%	93.19%
Stage 1 & Ensemble Balanced & ILO	93.86%	93.29%

Table 3: Sensitivity results for subset of considered models.

Business Insights from Uncertainty Quantification

To maximize the benefit of actionable conclusions, we also explore how the uncertainty of the model can be quanti-

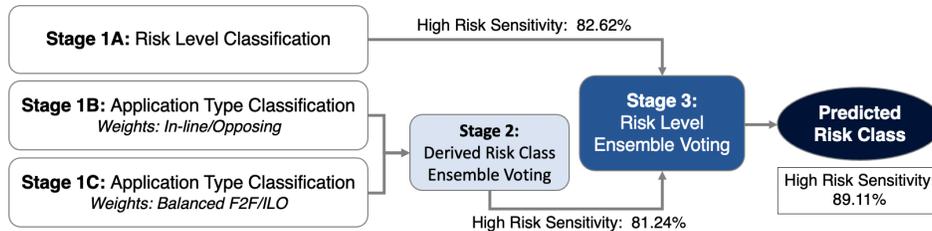


Figure 1: Chosen model architecture. Sensitivities shown are those produced by a single iteration on unknown data.

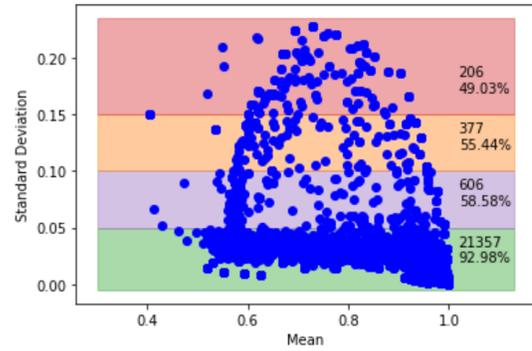


Figure 2: Standard deviation vs mean over 1000 models on bootstrapped datasets. Shading and annotations indicate count and accuracy of training records falling in shaded area.

fied, visualized and communicated to the representatives responsible for reaching out and to the customers themselves. As the purpose of this model is to identify COVID-unsafe workplaces, it is essential that outreach to prospective unsafe workspaces be supported by a robust understanding of how uncertain the model is with regard to each specific record (ie. workplace being classified). To evaluate this, we adopt a modified version of the uncertainty quantification approach presented by (Stracuzzi et al. 2018), where an ensemble of decision trees was trained (each on bootstrapped datasets) and then the mean and standard deviation of all of the decision trees' predictions on a single record were used by the authors as a proxy for the uncertainty of the prediction for that record. Our multi-stage model is more complex than a single decision tree, but we replicate the same approach with our already effectively trained model by creating 1000 separate versions of the multi-stage model on different bootstrapped datasets and, for each record, run all 1000 and compute the mean and standard deviation of the prediction probability. In order to do this, the model is modified to carry the probabilities through the different stages, rather than simply the predicted classes. The resulting mean and standard deviation for all records are illustrated in fig. 2, where we adopt the same visualization technique as in (Stracuzzi et al. 2018) and display the overall accuracy of all records in each horizontal colored area in the annotation on the right edge.

A more detailed visualization of high and low risk classes and correct and incorrect results is shown in fig. 3, illustrating that significantly more records predicted correctly have lower standard deviations compared to records predicted incorrectly. A t-test shows a statistically significant difference

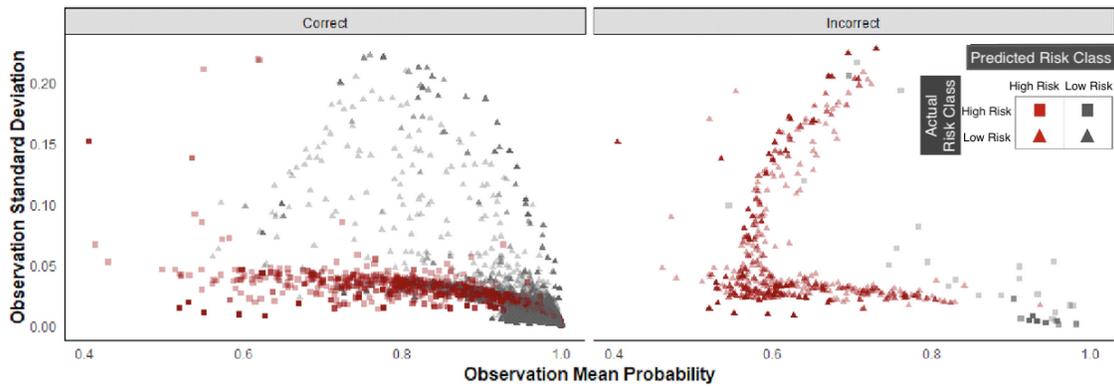


Figure 3: Standard deviation vs mean over 1000 models on bootstrapped datasets for each training record, separated by correct (left) and incorrect (right) predictions.

in standard deviation of correct and incorrect classifications of records predicted as high risk (with $p < 0.001$), indicating the standard deviation is a proxy for the reliability of a prediction.

Considering only the records with high risk classification, an even more clear pattern emerges with distinct sections of the plot. As illustrated in fig. 4, for high risk predictions, the results can be split into four quadrants with varying degrees of reliability of the predictions. Records in the lower right quadrant are extremely reliable (93.97% correct), lower left quadrant are more variable (21.58% correct), upper left quadrant are highly unreliable (6.43% correct), and upper right quadrant are extremely unreliable (all incorrect). This approach allows each individual observation to be assessed for the reliability of its prediction.

Performing the same process on a far larger set of “real world data” (data with unknown true classes), shows the same data spread pattern holds for the records predicted as high risk, indicating the quadrant analysis is generalizable. This is crucial, as it allows prioritizing which customers to reach out to first and provides an explainable understanding of the confidence that they have a high-risk workspace.

Conclusions

We have presented a multi-stage model capable of detecting workplaces at high risk for COVID-19 transmission by leveraging historical sales data and product information. The model is effective at classifying high- and low-risk workspaces, with a high high-risk sensitivity, ensuring that the vast majority of high-risk workspaces are detected. We have illustrated how existing approaches to uncertainty quantification in machine learning can be applied to create actionable, prioritized insights from the results for which the confidence in the prediction can be relatively easily articulated and understood.

This work will soon be deployed to identify and reach out to customers to facilitate a safe return to the office, and the leads generated from this analysis will be critical to prioritize “who to reach out to”. Results from the current model have led to the identification of close to 1.9 billion USD in office space investment that needs to be retrofitted in order to make it safe for organizations to re-enter the workspace. This work is expected to lead to millions of dollars in incremental revenue for the organization, but most importantly, it is expected to be a direct contributor in supporting organizations to reopen their office spaces safely.

References

- Hodari, J. 2015. Is working remotely sapping your creativity. *Harvard Business Review*.
- Nielsen, C. 2016. How microsoft used an office move to boost collaboration. *Harvard Business Review*.
- Pal, R.; Sekh, A. A.; Kar, S.; and Prasad, D. K. 2020. Neural network based country wise risk prediction of COVID-19. *Applied Sciences* 10(18):6448.
- Stracuzzi, D. J.; Darling, M. C.; Peterson, M. G.; and Chen, M. G. 2018. Quantifying uncertainty to improve decision making in machine learning. Technical Report SAND2018-11166 669595, Sandia National Lab.
- Wang, S.; Zha, Y.; Li, W.; Wu, Q.; Li, X.; Niu, M.; Wang, M.; Qiu, X.; Li, H.; Yu, H.; Gong, W.; Bai, Y.; Li, L.; Zhu, Y.; Wang, L.; and Tian, J. 2020. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *European Respiratory Journal* 56(2):2000775.

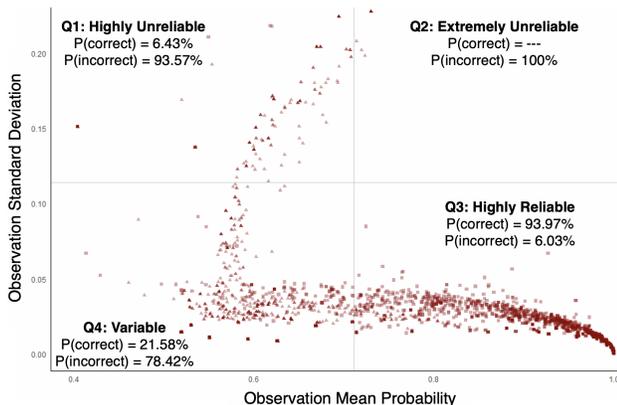


Figure 4: Standard deviation vs mean for training records classified as high risk. Squares and triangles represent correctly and incorrectly classified records, respectively.