MEF UNIVERSITY

ONLINE CHECK-IN LIKELIHOOD OF HOTEL GUESTS

Capstone Project

Yusuf Tiryaki

İSTANBUL, 2017

MEF UNIVERSITY

ONLINE CHECK-IN LIKELIHOOD OF HOTEL GUESTS

Capstone Project

Yusuf Tiryaki

Advisor: Prof. Dr. Özgür Özlük

İSTANBUL, 2017

MEF UNIVERSITY

Name of the project: Online Check-in Likelihood of Hotel Guests
Name/Last Name of the Student: Yusuf Tiryaki
Date of Thesis Defense: 11/09/2017

I hereby state that the graduation project prepared by Yusuf Tiryaki has been completed under my supervision. I accept this work as a "Graduation Project".

11/09/2017
Prof. Dr. Özgür Özlük

I hereby state that I have examined this graduation project by Yusuf Tiryaki which is accepted by his supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

11/09/2017

Director
of
Big Data Analytics Program

We hereby state that we have held the graduation examination of Yusuf Tiryaki and agree that the student has satisfied all requirements.

THE EXAMINATION COMMITTEE

Committee Member	Signature
1. Prof. Dr. Özgür Özlük	………………………..
2. …………………………..	………………………..

Academic Honesty Pledge

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

Yusuf Tiryaki 11/09/2017

EXECUTIVE SUMMARY

ONLINE CHECK-IN LIKELIHOOD OF HOTEL GUESTS

Yusuf Tiryaki

Advisor: Prof. Dr. Özgür Özlük

SEPTEMBER, 2017, 23 pages

Hotel operators benefit from current technological developments in order to provide the best experience for their guests to stay. In the case of an enterprise which providing guest hospitality service, the flow is composed of 4 steps: booking, check-in, accommodating and check-out. Online reservation systems have been in use for a long time and are services that offer room reservations for date ranges that guests will stay with. Online check-in applications are a new type of service that has just begun to be implemented in the hospitality sector. The advanced online hotel check-in systems enable users to save time by creating an entry log on the internet, specifying floor and room selection, assigning additional services, notifying the check-in time during the process, and reducing waiting times for hotel help desk during check-in.

In the online check-in forecasting process, a data analytics application was implemented that computes the score of the user's proximity to online check-in after the booking step and the booking information was obtained. The score calculation process uses statistical learning algorithms. Within the scope of the study, the guests were classified according to closeness to service reception with Random Forest and DNN(Deep Neural Networks) methods using a dataset in which the guests had hotel booking and provided online information. The trained model for classification was presented as a web service to return the likelihood score of new booking guests.

Key Words: hospitality, hotel online check-in, likelihood to buy, decision tree classification, deep neural network classification

ÖZET

OTEL MÜŞTERİLERİNİN ONLİNE CHECK-İN KULLANIMINI TAHMİNLEME

Yusuf Tiryaki

Tez Danışmanı: Prof. Dr. Özgür Özlük

EYLÜL, 2017, 23 sayfa

Otel işletmeleri konaklayacak misafirlerine en iyi deneyimi yaşatmak için güncel teknolojik gelişmelerden yüksek oranda faydalanmaktadır. Misafir ağırlama hizmeti sunan bir işletmede konaklama hizmeti sunumu özet olarak rezervasyon, giriş(check-in), konaklama ve ayrılış(check-out) adımlarından oluşmaktadır. Online rezervasyon sistemleri uzun zamandır kullanılmakta olup kullanıcılara konaklayacakları tarih aralığı için oda rezerve hizmetini sunan servislerdir. Online giriş(check-in) uygulamaları otelcilik sektörü için yeni uygulanmaya başlayan henüz gelişme aşamasında bir hizmet türüdür. Gelişmiş online otel check-in sistemleri internet ağı üzerinden giriş kaydı oluşturma, bu işlem sırasında kat ve oda seçimini belirleyebilme, ek hizmetleri atama, giriş saatinin bildirilmesi gibi işlemleri sağlayarak kullanıcıların zamandan tasarruf etmesini sağlamakta, giriş işlemleri sırasında otel yardım masalarındaki bekleme sürelerini kısaltmaktadır.

Online check-in kullanımı tahmin etme işlemi çalışmasında, rezervasyon adımı sonrası devreye giren ve rezervasyon bilgileri elde edilen kullanıcının online check-in yapmaya yakınlık skorunu hesaplayan bir veri analitiği uygulaması gerçeklenmiştir. Skor hesaplama işlemi istatistiksel öğrenme algoritmalarını kullanır. Çalışma kapsamında otele rezervasyon yapmış ve online check-in hizmeti sunumu yapılmış misafir bilgilerinin yer aldığı bir veri seti kullanılarak Random Forest ve DNN yöntemleri ile misafirler hizmet alımına göre sınıflandırılmıştır. Sınıflandırma için eğitilen model rezervasyon yapan yeni misafirlerin hizmet alımına yakınlık skorunu geri döndürmek üzere web servis olarak sunulmuştur.

Anahtar Kelimeler: otelcilik, otel online check-in, hizmet alma yakınlığı, ağaç bazlı sınıflandırma algoritmaları, derin öğrenme ağları ile sınıflandırma

TABLE OF CONTENTS

.

1. INTRODUCTION

The general changes in society have led to new expectations that redefine what needs to be competitive to satisfy guests and win loyalty. It has been the countless milestones celebrated in hospitality throughout the history of hotel technology, such as electricity, hotel phone, in-room radio, now the standard hotel room TV, and now it's been incorporated into the end of mobile technology.

The ongoing evolution of information technology has an important influence on the hospitality industry. The widespread use of the Internet has created many conditions that change the game. As a result, many hotels in the 21st century had to make important adjustments to remain solvent and permanent.

Nowadays, one of the services that hotels add to technology stacks to increase the quality of service is the online services. Online booking and online check-in procedures, which are reflected as time-saving and satisfaction increase for guests, are being offered by many hotel chains as a basic service.

1.1. Hospitality Online Reservation

Online hotel reservations are a popular method for booking hotel rooms. Travelers can book rooms on a computer by using online security to protect their privacy and financial information and by using several online travel agents to compare prices and facilities at different hotels (Barth & Walsh, 2008).

Prior to the Internet, travelers could write, telephone the hotel directly, or use a travel agent to make a reservation. Nowadays, online travel agents have pictures of hotels and rooms, information on prices and deals, and even information on local resorts. Many also allow reviews of the traveler to be recorded with the online travel agent.

1.2. Hospitality Online Check-in

Mobile and web check-in enhances guest options and overall experience at participating hotels by reducing queuing times and allowing guests to interact with the hotel at a time and place convenient them.

A fully automated online check-in solution enables the guest to manage all mundane processes themselves, saving precious time and avoiding delays. Upon arrival at the hotel, the guest is able quickly to collect their key from a key dispenser in reception and gain access to their room. A fully automated online system can also contain an online check-out service. For check-out, the guest has the choice managing their exit from the hotel in a way that is convenient for them also. For example, guest can check their bill, add items from the minibar, pay their bill and leave avoiding the risk delays manual checking out that may cause. Figure 1 is an illustration of automated hotel system.

Figure 1 – Automated Hotel

1.2.1. Advantages of Online Check-in Forecasting

The online check-in closeness score of the guest who completes the reservation process provides a very strong insight into the planning of the marketing approach to the guest. The closeness score can be used as a key feature in determining the channel to be reached to the guest, as well as the advanced check-in options to be presented during communication and the up-selling items to be offered.

A check-in closeness score is one of the most important inputs to the calculation of reservation cancellation probabilities. The hotel has the potential to be used as a decision support system by these units, as it can be useful for income management units in resource planning and price setting stages in estimating occupancy. Figure 2 is a general overview of how hotel guest online check-in usage forecasting system is integrated with other hospitality management components.

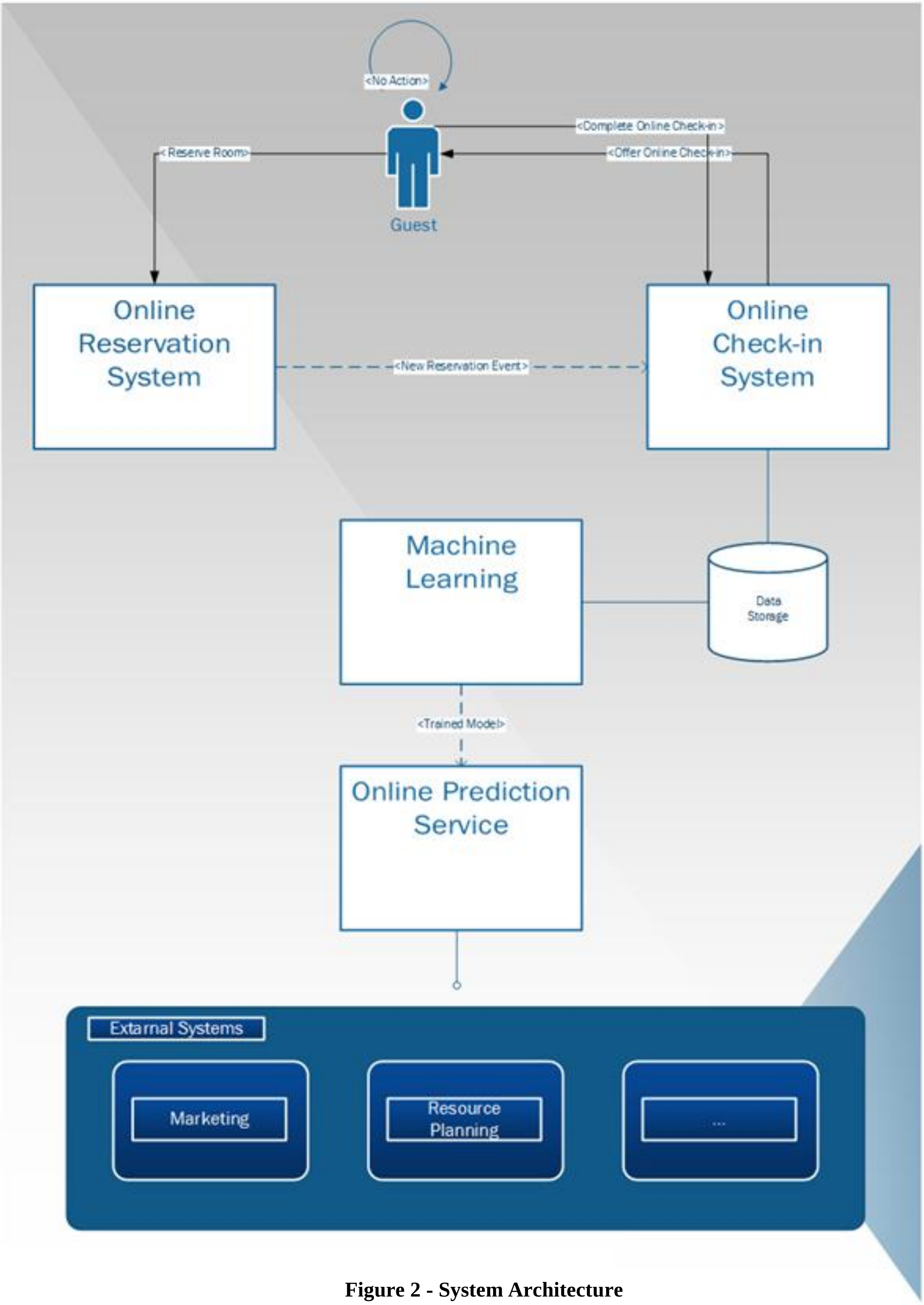


Figure 2 - System Architecture

2. APPLICATION

Two kinds of models were developed using the decision tree (Random Forest) and artificial neural network (DNN) while the hotel guest online check-in proximity estimation system was developed. The developed models were tested with samples that were never seen by the system and statistical consistency was examined. As a result of the model development and evaluation stages, the more successful DNN model has been used in the web service module to provide online prediction service.

2.1. About Data

The data used during the study were compiled from an online check-in service database, which is currently piloted. This database contains the details of all bookings made by the hotel, the specific demographic information of the guests and whether they use the online check-in system. The data are for a single hotel and include transactions made between August 2016 and April 2017. The dataset contains 2251 labeled rows with 18 attributes. The dependent attribute of the dataset is "*OnlineCheckin"* with the type binary(0,1). Most of the independent attributes introduce details of booking: *"BookingDate", "ArrivalDate", "Nights", "RoomType", "AdultCount"* features of booking is included in the dataset and used by the model. *"EmailDomain", "Gender", "CountryCode"* are guest specific attributes and they also used by the model to predict guests online check-in behavior. For detailed information of the dataset, please refer to APPENDIX A.

2.2. Methodology

Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems. Cases such as the digit recognition example, in which the aim is to assign each input vector to one of a finite number of discrete categories, are called classification problems (Mitchell, 1997).

Since our dataset includes information that whether guests do online check-in, online check-in proximity prediction can be turned into a two-class supervised learning problem. Using classification algorithms, I predicted which class the guest more close to and evaluated the performance values of the obtained results according to the benchmarking criteria in the literature and assessed whether they were within acceptable limits.

2.2.1. Hotel Guest Online Check-in Likelihood

The main objective of this study is to show that online check-in likelihood of hotel guests is a learnable problem by machine learning methodologies. In the terminology of statistics if define the problem: guests have two categories as *OnlineCheckin:YES* and *OnlineCheckin:NO*. In this case, each input vector may belong to one of two categories. The probability that a new sample belongs to a category will be that class distance, and the sum of these probabilities must equal 1.

$$p(OnlineCheckin\colon YES) = 1 - p(OnlineCheckin\colon NO)$$

Every observed response is in set of binary labeled examples:

$$y_i \in \{OnlineCheckin\colon NO, OnlineCheckin\colon YES\} \equiv y_i \in \{0,1\}$$

The main objective function to minimize is:

$$\sum_{i=1}^{N} y_i \log \widehat{y_i} + (1 - y_i) \log(1 - \widehat{y_i})$$

2.2.2. Feature Selection

Features of dataset were visualized with data visualization tools to understand the explanatory properties of the arguments in the data set.

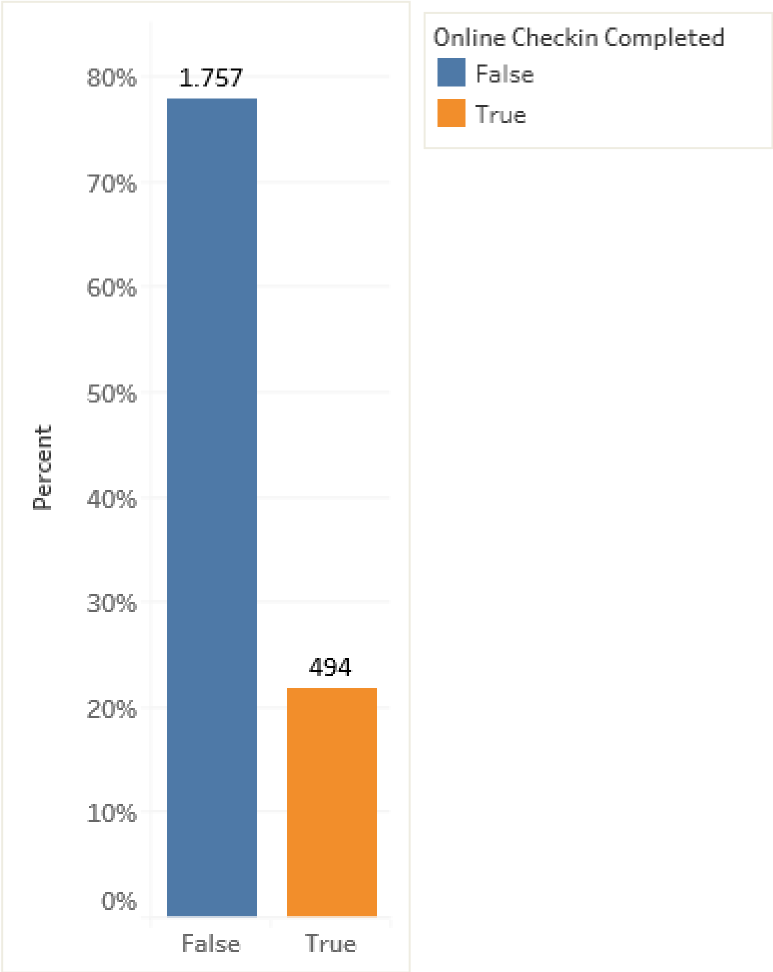


Figure 3 - Online Check-in Class Distribution

As seen in Figure 3 - Online Check-in Class Distribution we can’t talk about an imbalance distribution of classes as they are 76% *OnlineCheckin:NO* labeled and 24% are *OnlineCheckin:YES* labeled.

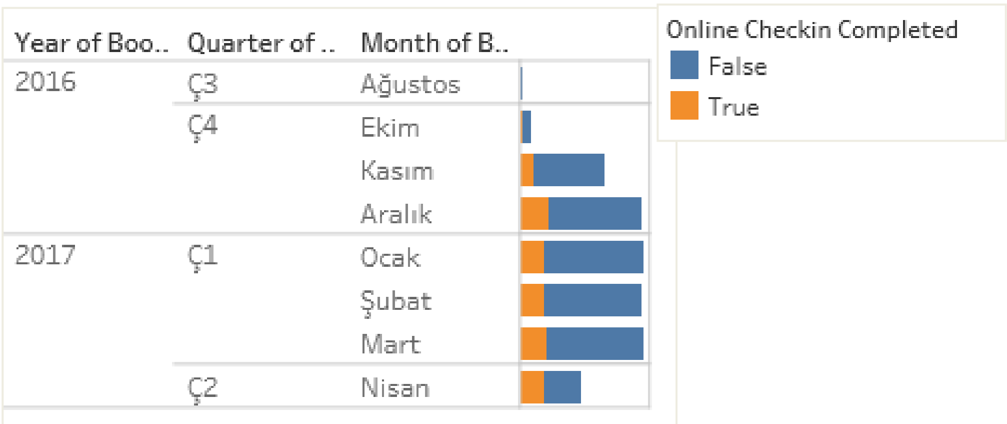


Figure 4 - Online Check-in By Date

Figure 4 - Online Check-in By Date has shown that any booking date based categorization has not a powerful effect on classification as the distribution of classes are similar for all booking months.

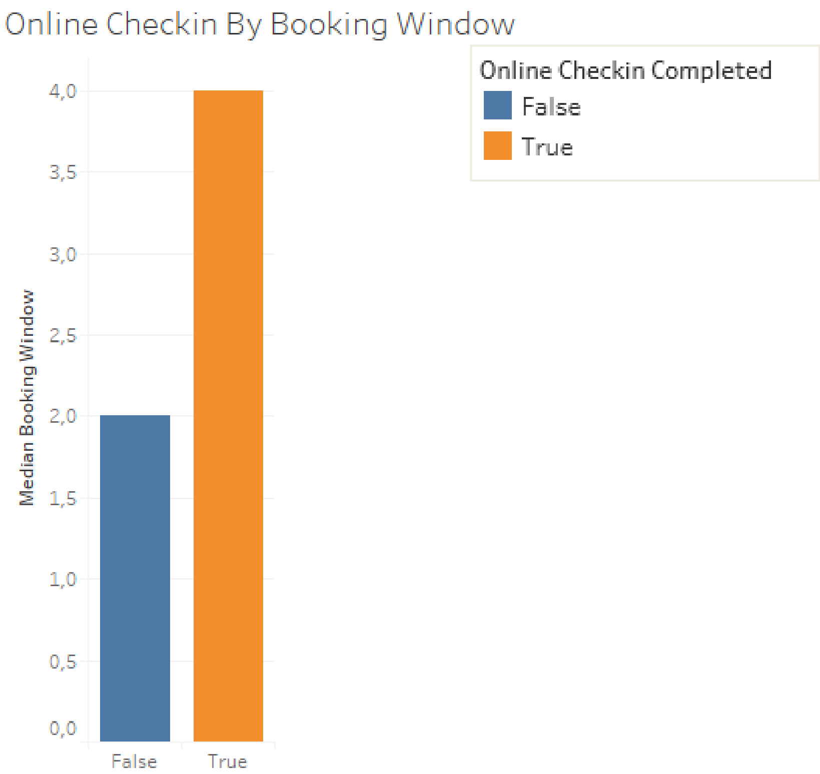


Figure 5 - Online Check-in By Booking Window

The time window between the reservation date and the date of arrival was considered to have the explanatory power and this feature was created as name "*BookingWindow"* and visualized. As can be seen in Figure 5 - Online Check-in By Booking Window, online check-in completed guest have a higher booking window value than not online checked-in guests therefore new feature added to the model.

The reservation date and the date of arrival information are converted to information of the day of the year so that they can be used in the classification algorithm.

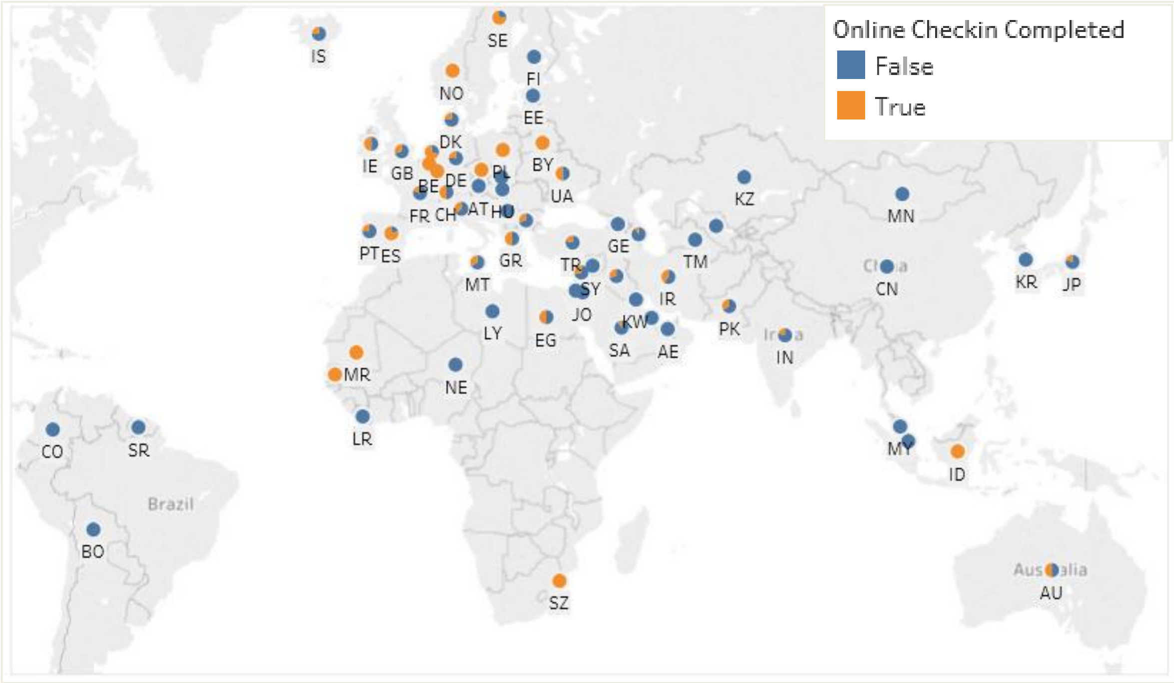


Figure 6 - Online Check-in By Country

Figure 6 - Online Check-in By Country has shown that guests from different countries have different online check-in behavior. The country attribute is a good candidate to sharpen purity of the classes.

2.2.3. Random Forest Model

RF classification uses an ensemble of unpruned decision trees, each grown using a bootstrap sample of the training data, and randomly selected subsets of predictor variables as candidates for splitting tree nodes. The RF regression prediction for a new observation x ($\hat{f}_{rf}^{B}(x)$) is made by averaging the output of the ensemble of B trees $\{T(x,\Psi_b)\}_1^B$ as:

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x,\Psi_b),$$

where Ψ b characterizes the both RF tree in terms of split variables, cutpoints at each node, and terminal node values (Hastie, Tibshirani, & Friedman, 2009).

I implemented RF in the R package randomForest with decision trees as base learners (Liaw & Wiener, 2002). Following various recommendations, I evaluated different combinations of the values of the number of trees to grow, ntree = {50, 100, 1000}, the number of SNPs randomly selected at each tree node, mtry = {0.5, 1, 2} × the default value of mtry of square root of sample size for classification, and the minimum size of terminal nodes of trees, below which no split is attempted, nodesize = 1. The parameter configuration with the highest prediction accuracy was ntree =100, mtry = 8 and nodesize =1.

The values of true positive rate and false positive rate are often represented diagrammatically by a ROC graph. Joining the points on a ROC graph to form a ROC curve can often give insight into the best way of tuning a classifier. A Euclidean distance measure of the difference between a given classifier and the performance of a hypothetical perfect classifier is described (Bramer, 2013). The area under ROC curve is often used as a measure of the quality of the classification models. A random classifier has an area under the curve of 0.5, while AUC for a perfect classifier is equal to 1. In practice, most of the classification models have an AUC between 0.5 and 1. A ROC curve demonstrate:

- It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
- The closer the curve follows the upper-left border of the ROC space, the more accurate the test.
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
- The slope of the tangent line at a cutpoint gives the likelihood ratio (LR) for that value of the test.
- The area under the curve is a measure of accuracy.

As evaluation metric, I have used area under curve(AUC) of ROC graph. Table 1 is an AUC score quality identification mapping which is widely used in industry (Tape, 2017).

AUC	**Identifier**
.90 - 1	Excellent
.80 - .90	Good
.70 - .80	Fair
.60 - .70	Poor
.50 - .60	Fail

Table 1 - AUC Standards

Full list of the model training parameters are represented at Table 2 - Model Parameters

Parameter	Value	Description
response_column	OnlineCheckin	Response variable column.
ntrees	100	Number of trees.
mtries	8	Number of variables randomly sampled as candidates at each split. If set to -1, defaults to sqrt{p} for classification and p/3 for regression (where p is the # of predictors
max_depth	20	Maximum tree depth.
min_rows	1	Fewest allowed (weighted) observations in a leaf.
nbins	20	For numerical columns (real/int), build a histogram of (at least) this many bins, then split at the best point.
sample_rate	0.632	Row sample rate per tree (from 0.0 to 1.0)

Table 2 - Model Parameters

As dataset is split with the 0.75 proportion for training 0.10 for validation and 0.15 for test, Random Forest model has trained with the 0.8 of data and 0.8498 AUC(Figure 7 – RF Training ROC Curve) has retrieved. As mentioned in Table 1 - AUC Standards it is the in the class of a Good model. As seen in Table 4 - RF Training Confusion Matrix error rate for *OnlineCheckin:YES* cases 0.49 which is nearly a random model. However the model is successful in the field of predicting the likelihood of actual class described by AUC score. The trained model also has introduced a variable importance diagram(Figure 8 – RF Variable Importance) which verifies consistency of variable selection decisions during feature selection phase.

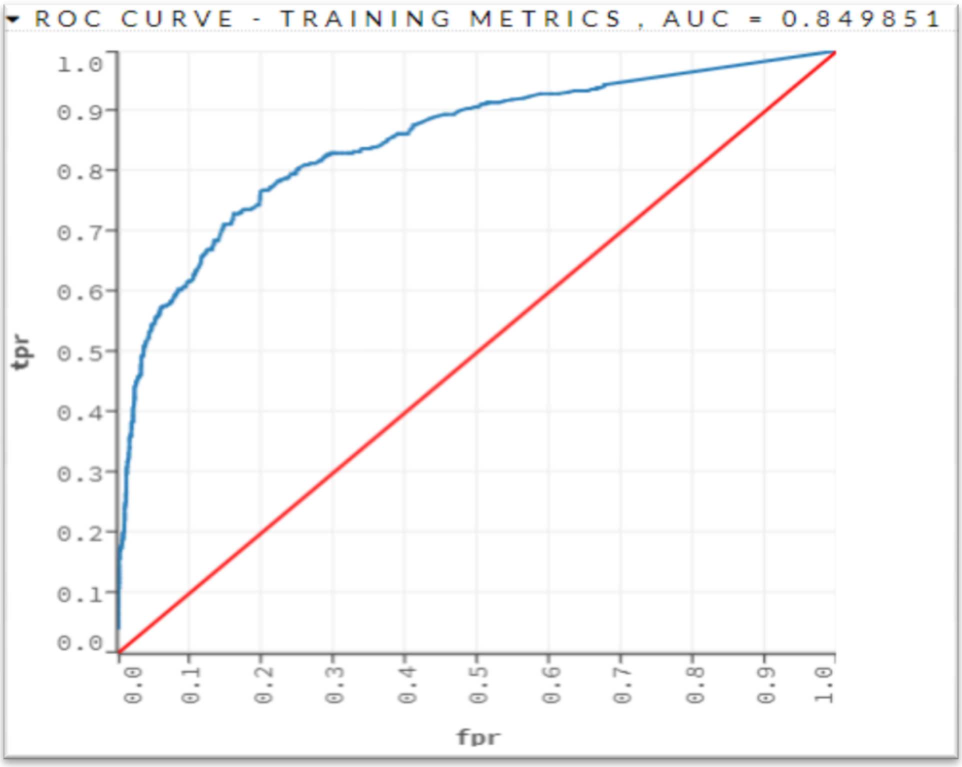


Figure 7 – RF Training ROC Curve

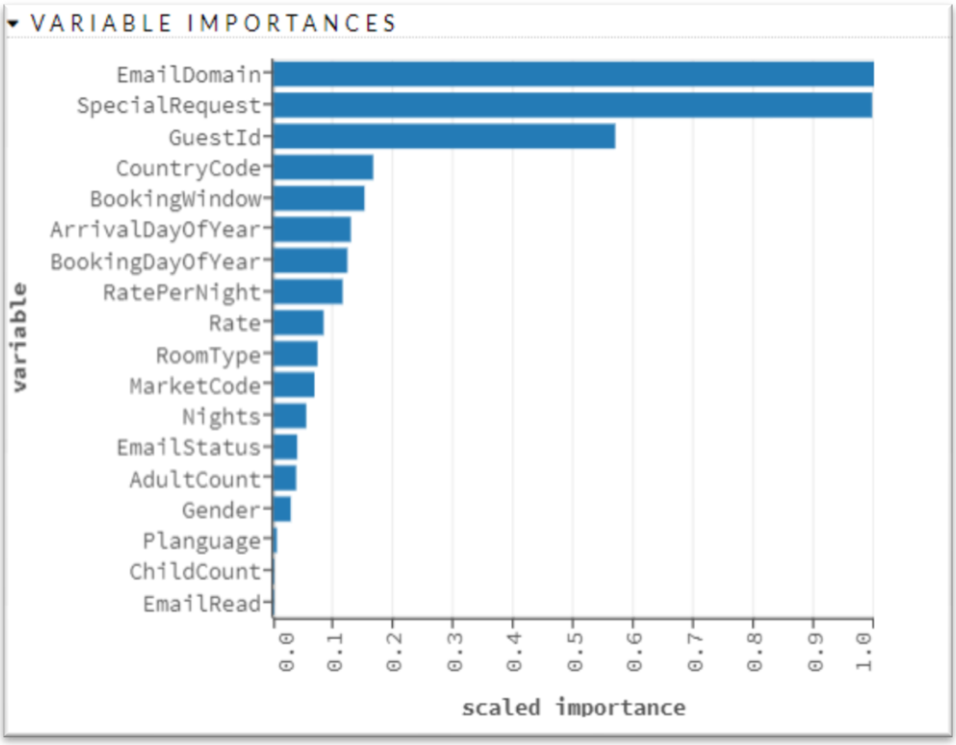


Figure 8 – RF Variable Importance

MSE	0.110238
RMSE	0.332021
r2	0.369165
logloss	0.774605
AUC	0.849851
Gini	0.699702
mean_per_class_error	0.242846

Table 3 – RF Training Metrics

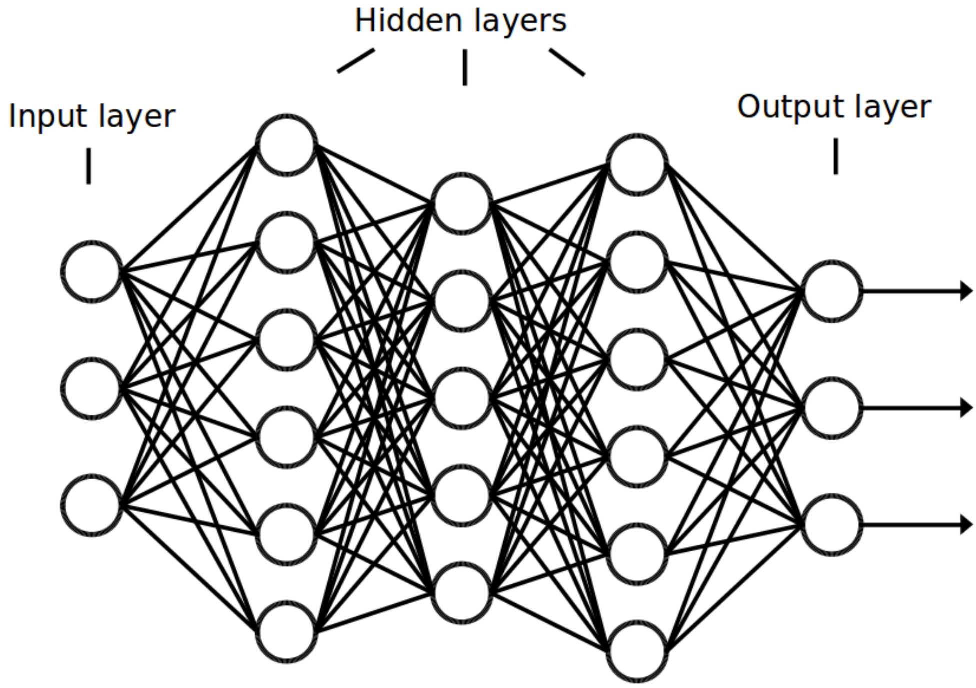
Actual/Predicted	**NO**	**YES**	**Error**	**Rate**
OnlineCheckin				
NO	1343	50	0.0359	50 / 1393
YES	199	207	0.4901	199 / 406
Total	**1542**	**257**	**0.1384**	**249 / 1799**

Table 4 - RF Training Confusion Matrix

2.2.4. Deep Artificial Neural Network Model

The term 'neural network' has its origins in attempts to find mathematical representations of information processing in biological systems (Rosenblatt, 1962).

An ANN is based on a collection of connected units called artificial neurons, (analogous to axons in a biological brain). Each connection (synapse) between neurons can transmit a signal to another neuron. The receiving (postsynaptic) neuron can process the signal(s) and then signal downstream neurons connected to it. Neurons may have state, generally represented by real numbers, typically between 0 and 1. Neurons and synapses may also have a weight that varies as learning proceeds, which can increase or decrease the strength of the signal that it sends downstream. Further, they may have a threshold such that only if the aggregate signal is below (or above) that level is the downstream signal sent. Figure 9 is a graphical representation of artificial neural networks (Jordan, 1999).



In my study I used a 7-layer deep network with ReLU activation function, trained model and measure the success of learning to divide into correct outcome classes. I used the AUC ROC graph metric as in the Random Forest model as a performance criterion. I have applied hidden nodes selection in hidden layer guidelines as described by (Panchal & Panchal, 2014). In order to avoid the model overfitting problem during the model training, I applied the L2 regularization which prevents very large values in the weight vector. In L2 regularization, you modify the error function you use during training to include an additional term that adds a fraction of the sum of the squared values of the weights. So larger weight values lead to larger error, and therefore the training algorithm favors and generates small weight values (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). Table 5 shows the set of parameters used for model training.

Parameter	***Value***	***Description***
response_column	OnlineCheckIn	Response variable column.
hidden	18, 48, 66, 84, 102, 120, 138, 156	Hidden layer sizes
epochs	50	How many times the dataset should be iterated (streamed), can be fractional.

Figure 9- Artificial Neural Network

l2	0.0005	L2 regularization (can add stability and improve generalization, causes many weights to be small.
adaptive_rate	true	Adaptive learning rate.
rho	0.99	Adaptive learning rate time decay factor (similarity to prior updates).
epsilon	1e-8	Adaptive learning rate smoothing factor (to avoid divisions by zero and allow progress).
rate	0.005	Learning rate (higher => less stable, lower => slower convergence).
rate_annealing	0.000001	Learning rate annealing: rate / (1 + rate_annealing * samples).
rate_decay	1	Learning rate decay factor between layers (N-th layer: rate * rate_decay ^ (n - 1).
momentum_start	0	Initial momentum at the beginning of training (try 0.5).
momentum_ramp	1000000	Number of training samples for which momentum increases.
momentum_stable	0	Final momentum after the ramp is over (try 0.99).
nesterov_accelerated_gradient	true	Use Nesterov accelerated gradient (recommended).
input_dropout_ratio	0	Input layer dropout ratio (can improve generalization, try 0.1 or 0.2).
hidden_dropout_ratios		Hidden layer dropout ratios (can improve generalization), specify one value per hidden layer, defaults to 0.5.
l1	0	L1 regularization (can add stability and improve generalization, causes many weights to become 0).
l2	0.0005	L2 regularization (can add stability and improve generalization, causes many weights to be small.
max_w2	3.4028235e+38	Constraint for squared sum of incoming weights per unit (e.g. for Rectifier).

Table 5 - DNN Training Parameters

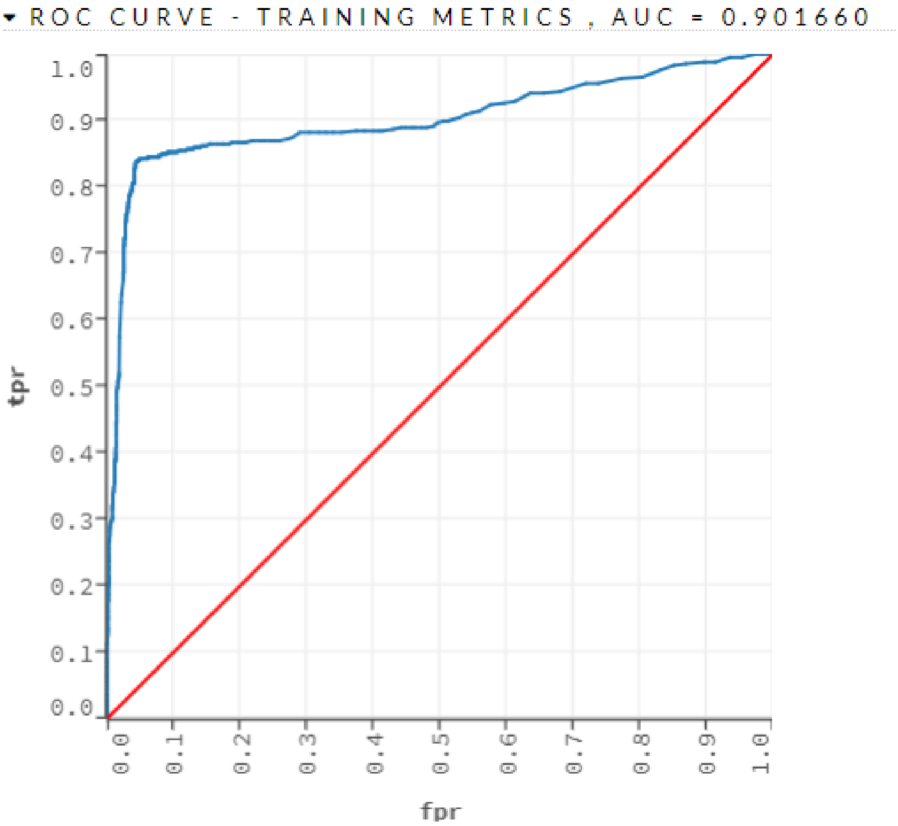


Figure 10 - DNN AUC

MSE	0.073936
RMSE	0.271911
r2	0.576903
logloss	0.282347
AUC	0.901660
Gini	0.803320
mean_per_class_error	0.102817

Table 6 - DNN Training Metrics

Actual/Predicted OnlineCheckin	**0**	**1**	**Error**	**Rate**
0	1334	59	0.0424	59 / 1393

1	67	339	0.1650	67 / 406
Total	**1401**	**398**	**0.0700**	**126 / 1799**

Table 7 - DNN Training Confusion Matrix

3. RESULTS

15% of the dataset was allocated as a test set and this data was never used during the training phase. Performance results were obtained by giving these test data as input to the trained models. The results of the DNN model showed better results both during training and during the test, and it was able to explain at an acceptable level the likelihood of online check-in usage of hotel guests.

3.1. Random Forest Model Scores

As represented in Figure 11 - RF Validation AUC and Table 8 - RF Validation Metrics validation scores of Random Forest model is a good candidate to predict the likelihood of online hotel check-in. AUC score is 0.81 which is appreciated as a Good model by the industrial standards. However, it is not as skilled as DNN to convergence the soft edges of the domain.

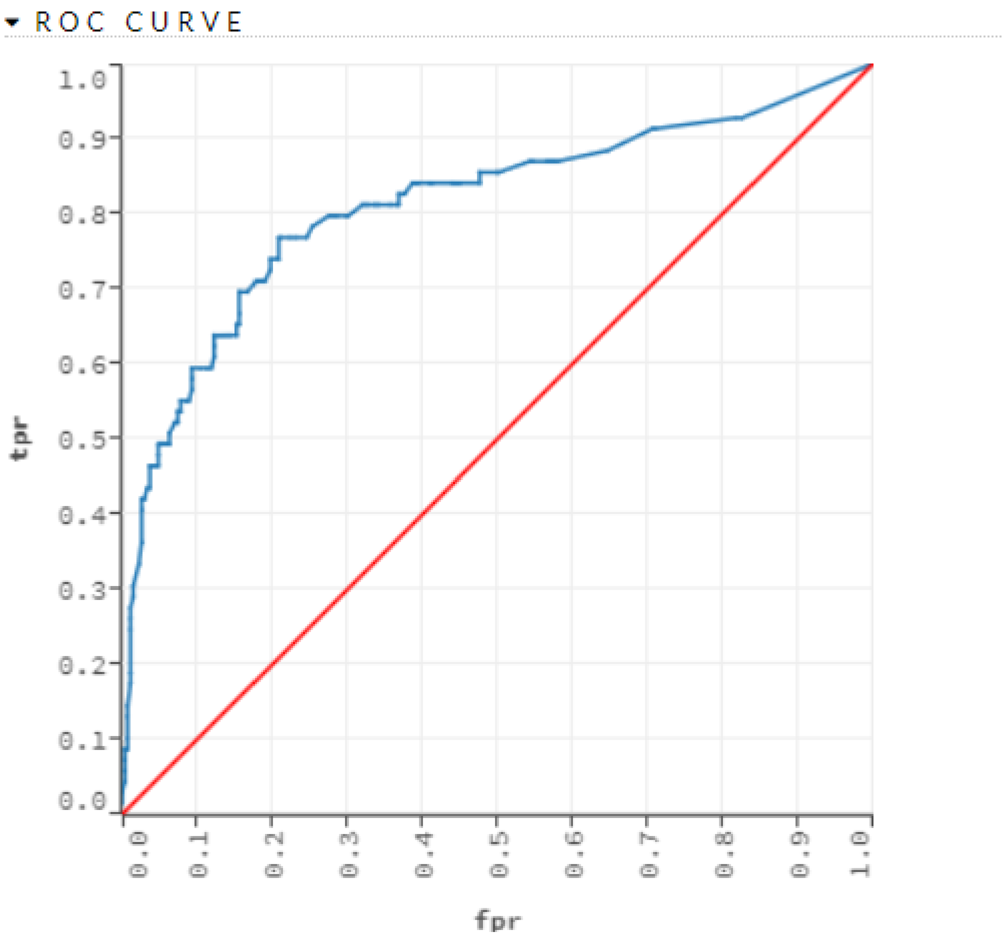


Figure 11 - RF Validation AUC

MSE	0.115293
RMSE	0.339549
r2	0.290363
logloss	0.840988
AUC	0.812322
Gini	0.624643
mean_per_class_error	0.249367

Table 8 - RF Validation Metrics

Actual/Predicted OnlineCheckin	NO	YES	Error	Rate
NO	244	25	0.092937	25/269
YES	28	41	0.405797	28/69
Total	**272**	**66**	**0.156805**	**53/338**

Table 9 - RF Validation Confusion Matrix

3.2. DNN Model Scores

The AUC score on the test set of the DNN model was 0.84 (Table 10 - DNN Validation Scores). When evaluated together with confusion matrix metrics (Table 11 - DNN Validation Confusion Matrix), it was concluded that the performance in predicting was sufficiently accurate and robust to be used in decision support systems.

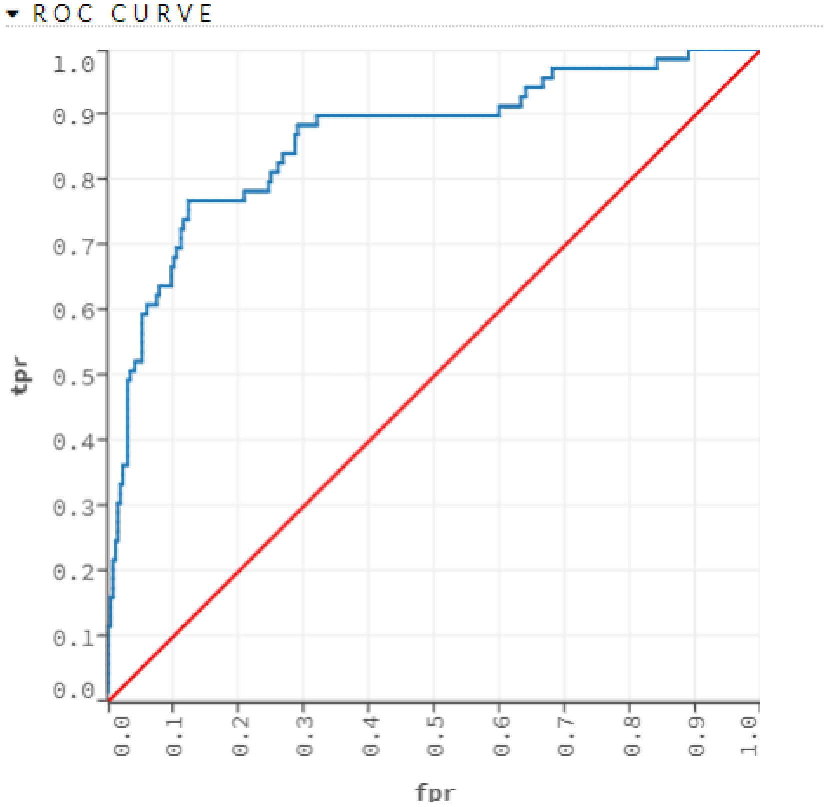


Figure 12 - DNN Validation AUC

MSE	***0.105017***
RMSE	***0.324063***
r2	***0.353617***
logloss	***0.378591***
AUC	***0.864393***
Gini	***0.728786***
mean_per_class_error	***0.177280***

Table 10 - DNN Validation Scores

Actual/Predicted OnlineCheckin	**NO**	**YES**	**Error**	**Rate**
NO	236	33	0.122677	33/269
YES	16	53	0.231884	16/69
Total	**252**	**86**	**0.144970**	**49/338**

Table 11 - DNN Validation Confusion Matrix

4. SUMMARY

During my study, I have constructed an experiment to predict online check-in likelihood of hotel guests. I have used reservation details and guest's demographic details as independent variables and tried various machine learning models for minimizing prediction errors. I have shown that a fine-tuned Deep Neural Network model can predict the likelihood of online check-in in hotels with a Good labeled quality by industrial standards.

4.1. Predictive Analytics and Hospitality

This study anticipates the online check-in choices of hotel guests and allows the hotel industry's resource planning department to plan the resources to check-in very precisely. The marketing department has been able to use this information to communicate with the guest during the online check-in phase and to support the decision to make the side products available.

4.2. Further Works and Ethical Aspects

Online check-in Likelihood estimating success can be encouraging for many processes that process similar scenarios. Again in the hospitality industry, there is a potential to predict other behaviors of the guests. For example, it seems possible to estimate the closeness of guests to use other services such as tour, massage parlor, laundry. Apart from that, it is also possible to present the application of the online check-in approximation to online reservations such as the aviation sector.

Many organizations and sections of world society are concerned with ethical behavior. (Lee & Tsang, 2013) assert that ethics is an important challenge in the hospitality industry, and claim that the understanding of ethical perception and moral position of all stakeholders should be accentuated. Since hotel guests have to share a lot of private information with the ones they stay, the ethical principles of personal data are one of the most important issues. Although the results of this study do not produce an output that affects the hotel guests individually, the final output is special to each person. Third-party business lines that will use these outputs should observe the national and international laws on the use and sharing of private personal data.

APPENDIX A

Feature	Description	**Type**
GuestId	Unique identifier for guest who completed online booking.(GuestId is same for different bookings of same guest)	string
BookingDate	Booking date and time	datetime
ArrivalDate	Arrival date	datetime
Nights	Number of nights to stay	numeric
AdultCount	Number of guests to stay	numeric
ChildCount	Number of child guest to stay	numeric
RoomType	Selected room type to stay ● H1K ● R1K ● R1T ● R2K ● R3K	categorical
Rate	Paid price for whole stay	numeric
MarketCode	The marketing channel which booking completed ● Corporate Group ● Corporate Local ● FIT Discount ● FIT Rack ● Individual Leis ● Non Tour Group ● Packages	categorical
SpecialRequest	Special requests made by guest during booking ● HIGH ● KNG ● NS ● QUI ● SM	categorical
EmailDomain	Email domain of guests email	string
EmailStatus	Status of email which sent to guest to complete online check-in ● bounce	categorical

	• delivered • dropped • processed	
EmailRead	Email read status	binary
IsLogin	Data of if guest has logged in to online check-in system	binary
Gender	Gender of guest • M • F	categorical
Title	Title of guest	string
Planguage	Guest language choice to communicate • E • TR	categorical
CountryCode	Country of guest • AU • AZ • BE • BG • …..	categorical
OnlineCheckin	Is guest completed online check-in or not	binary

REFERENCES

Barth, J. E., & Walsh, J. (2008). An Empirical Approach to Developing Classification and Rating Schemes. *Journal of Hospitality & Leisure Marketing* , 15-29.

Bramer, M. (2013). *Principles of Data Mining.* London: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements Of Statistical Learning.* New York: Springer.

Jordan, M. I. (1999). *Learning in Graphical Models.* Cambridge: MIT Press.

Lee, L. Y., & Tsang, N. K. (2013). Perceptions of tourism and hotel management students on ethics in the workplace. *Journal of Teaching in Travel & Tourism* , 228-250.

Liaw, A., & Wiener, M. (2002, 2). Classification and regression by randomForest. *R News* , s. 18-22.

Mitchell, T. M. (1997). *Machine Learning.* New York: McGraw-Hill.

Panchal, F. S., & Panchal, M. (2014, November). Review on Methods of Selecting Number of. *International Journal of Computer Science and Mobile Computing* , s. 455-467.

Rosenblatt, F. (1962). *Principles of neurodynamics: perceptrons and the theory of brain mechanisms.* Michigan: Spartan Books.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014, June). A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* .

Tape, T. G. (2017, 1 1). *The Area Under an ROC Curve.* Retrieved 8 1, 2017, from University of Nebraska Medical Center: http://gim.unmc.edu/dxtests/roc3.htm

MEF UNIVERSITY

Name of the project: Online Check-in Likelihood of Hotel Guests
Name/Last Name of the Student: Yusuf Tiryaki
Date of Thesis Defense: 11/09/2017

I hereby state that the graduation project prepared by Yusuf Tiryaki has been completed under my supervision. I accept this work as a “Graduation Project”.

11/09/2017
Prof. Dr. Özgür Özlük

I hereby state that I have examined this graduation project by Yusuf Tiryaki which is accepted by his supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

11/09/2017

Director
of
Big Data Analytics Program

We hereby state that we have held the graduation examination of Yusuf Tiryaki and agree that the student has satisfied all requirements.

THE EXAMINATION COMMITTEE

Committee Member	Signature
1. Prof. Dr. Özgür Özlük	………………………..
2. ………………………..	………………………..

Academic Honesty Pledge

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

Yusuf Tiryaki 11/09/2017