MAPFRE | April 2020

**MAPFRE Digital Workplace Data Analytics**

**Authors:**

|  |  |  |  |
| --- | --- | --- | --- |
| *Weinan Zhi*  *<*[*wzhi@wpi.edu*](mailto:wzhi@wpi.edu)*>* | *Chuchen Dai*  *<*[*cdai2@wpi.edu*](mailto:cdai2@wpi.edu)*>* | *Tahrifa Tarannum Lisa*  *<*[*ttlisa@wpi.edu*](mailto:ttlisa@wpi.edu)*>* | *Omkar Kulkarni*  *<*[*okulkarni@wpi.edu*](mailto:okulkarni@wpi.edu)*>* |

**Abstract**

The white paper will provide information that will be significant in the Digital Workplace effort in 2020. It will portray how digital adoption and collaboration, meeting improvements, and digital communication can be expanded. The document will also provide advanced analytics and insights on Office 365 data on an overall basis as well as by personas similar to what is provided in Microsoft MyAnalytics on an individual basis.

**Key words**

Office 365, Time series analysis, Clustering, Persona, KPI, User, Activity, Survey

**Section 1:**

**1.1 Problem Statement**

The main point of the project is to boost productivity in the public workplace using predictive modelling and app usage analytics insights. The questions which are to be answered are how the tools available at disposal of employees at Mapfre are being used now and how can that be improved for better productivity, what are the factors affecting that, what are the key performance indices that reflect on the persona creation and finally how to make employees proficient using those personas.

**1.2 Project Objectives and Summary of Contributions**

* Complete Analysis of O365 Data using 2019 data given to build a baseline model and further analyze the 2020 data against that baseline.
* Optimize use of O365 data with the data compiled, reviewed and analyzed to identify Key Performance Indices (KPI’s)
* Recommend additional KPIs based on the analysis if needed.
* Design a model that can be used to measure employee’s adaption of new ways of working by persona creation and company division.

**Section 2:**

**Methodology**

Major parts of the project are: the Time series analysis and the Persona analysis.  Firstly, for the Time Series part, instead of using each field provided as target value, we designed indicators for each application, which consists of those provided fields, to be our target predicting values in order to evaluate the usage of each application. For each indicator specifically, ARIMA models are implemented as baseline models. However, ARIMA does not perform well on most of the indicators’ future trend prediction. Based on that, various more advanced models are implemented such as PROPHET, LSTM and RNN models to get ideal trend prediction.  Secondly, for the Personas part, the employee persona system will be composed of KPI results and summarized persona tags. We used 3 kinds of data from different sources to make the final result reasonable. These data are: office 365 activities that show employees’ activity patterns and trends, employee census data that include some basic information about employees, the survey results that we designed with the MAPFRE managers to know employees’ subjective ideas better.

**Section 3:**

**Survey Questionnaire**

In order to get detailed employee data tailored to KPI’s, a company wide survey has been conducted which has information about department name, employee id, skillset specific questions and how often they use specific softwares and what are their preferences about them and how they share the files company-wide.

**Section 4:**

**Time Series Analysis**

On the applications data provided, ARIMA models are the baseline models and RNN, LSTM and PROPHET models are done and the most accurate results are chosen for analysis. For this analysis, the given data is for a period of 180 days. We have divided the data into train and test by 18 days margin.

**Section 5:**

**Persona Analysis**

Based on the new Audit Log Data and Survey Questionnaire we will evaluate an employee on KPIs. These KPIs will show some relevant patterns of employees by their grades. Here is some of our initial KPIs:

* Active Level / Adaptation - mainly generated from: Office 365 Activity, Survey)
* Communication Skill - mainly generated from: Skype Activity, AuditLog data
* Working Efficiency - mainly generated from: Office365 Activity, AuditLog data, Census
* Cooperative Level - Mainly Generated from: Teams Activity, Skype

**Clustering Analysis**

In order to decide the persona employee system, various types of clustering have been performed. Using K-means clustering as a baseline model, Hierarchical clustering, DBScan clustering and SOM model have been implemented. Overall, we proposed four Clustering results. K-means results show the least informative Clustering results as expected. Our expectation on the Clustering results is the Neural Network Clustering will outperform all the other Clustering methods with enough depth and complexity. Due to the time limit of our project, we do not have enough time to discuss and design a sound architecture of Clustering Neural Network and tune the hyperparameters accordingly. We are left with a choice between Hierarchical Clustering results and DBScan results. Generally, these two results are showing similar level of dissimilarity between clusters. We suggest the DBScan model as by far the best results since it shows more characteristics on Teams activity which is the focus of our project and can be further utilized into GUI visualization.

**Conclusion**

The white paper illustrates the initiative of improving efficiency and productivity of the employees of MAPFRE Insurance Company by the analysis and preprocessing of the provided data of the four applications namely, Skype, SharePoint, Teams and Onedrive. In order to achieve this goal, the data has been used to create different machine models and compare the outputs and choose the models which are the best.

**References**

1[KERAS API](http://www.keras.io)

2*Key Performance Indicators for dummies* by Bernard Marr

3 “A General Coefficient of Similarity and Some of Its Properties” by J. C. Gower

4 [ARIMA Models for Time Series Forecasting](https://people.duke.edu/~rnau/411arim.htm)

5[Facebook PROPHET Python API](https://facebook.github.io/prophet/docs/quick_start.html#python-api)

6 [Microsoft O365 Activity Report Value Mappings](https://docs.microsoft.com/en-us/microsoft-365/admin/activity-reports/activity-reports?view=o365-worldwide)