



## High Availability of Datacenter in Karachi, Pakistan

Fahim Uz Zaman<sup>1</sup>, Mansoor Ahmed Khuhro<sup>2</sup>, Hafsa Bint E Akram<sup>3</sup>, Ayesha Kalhoro<sup>4</sup>

<sup>1</sup> Department of Computer Science, SMI University, Karachi, Pakistan, faheemuzzaman@hotmail.com

<sup>2</sup> Department of Computer Science, SMI University, Karachi, Pakistan, makhuhro@smiu.edu.pk

<sup>3</sup> Department of Computer Science, SMI University, Karachi, Pakistan, hafsaakram50@gmail.com

<sup>4</sup> Department of Computer Science, SMI University, Karachi, Pakistan, kalhoroayesha@gmail.com

### ABSTRACT

The datacenter is the core infrastructure of today's world. Every data center should have many resources and applications that are running for several decades or even more. Many failures happen in the physical datacenter in on-premises environments. In this research paper evaluating the datacenter by having the dataset provided by Premier Systems (Pvt.) Ltd. This dataset is having all the failures and datacenter related issues from Jan-2016 to Dec-2019 in Karachi, Pakistan. This research performed the Linear Regression via the Microsoft Azure Machine Learning Studio for the machine learning model. It would allow us to know which fault will more and which is not for the concern requirement. This experiment would have the Feature Engineering feature in Microsoft Azure Machine Learning Studio, which will automatically apply the filters required. After knowing which the central issue are related to any physical data center in Karachi. This research allows to handle the required precaution in datacenters of Karachi, Pakistan.

**Key words:** High Availability, Data Centers, Servers.

### 1. INTRODUCTION

In this paper, validating the on-premises datacenter's high-availability. For this, we will perform the Machine Learning algorithm onto the data, which is provided by Premier Systems (Pvt.) Ltd. Premier Systems is a well-known enterprise-level company providing support and after-sale support to different vendors like Hewlett Packard, Juniper, Cisco, Microsoft, Vertiv, ABC, VMware, and Oracle, etc. They would provide the appliances and the supported weather; it is the software or hardware in nature. In this research, we will use the data, which is the property of Premier Systems (Pvt.) Ltd. This data is having the faults and their respective medication in the period of Jan-2016 to Dec-2019. During the

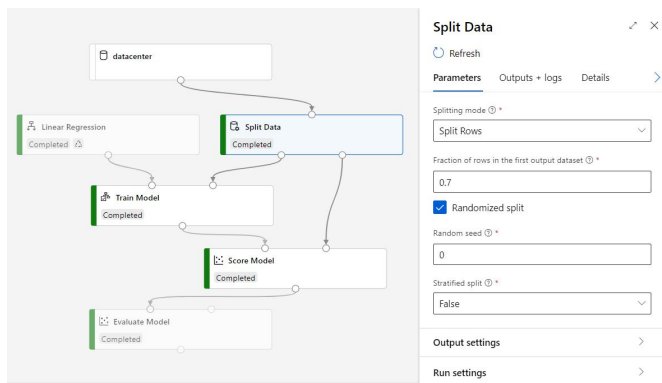
experiment would be using Machine Learning capabilities with Microsoft Machine Learning Studio. In this research, performing the Linear Regression - model to predict the most common failure things in the physical datacenter in Karachi. The dataset which is used specific to the Karachi, Pakistan region. No, other towns are included in this dataset.

### 2. RELATED WORK

Linear Regression in the best Machine Learning Model. Which is sincerely express by this research paper [4]. In this paper, you would see that every Machine Learning Model is now utilizing Linear Regression procedures' technicalities. During the experiment applied the Microsoft Azure Machine Learning capabilities by using the Microsoft Azure Machine Learning Studio. For this, significant research in this paper [1]. Also have another paper that works as that will be performed in our research paper [2]. As serve the Text Base classification for that, you would look onto this [3] paper. During research having a versatile group of studies already of the same nature. But, in different dataset via the Microsoft Azure Machine Learning Capabilities. In this experiment performed the same workflow structure, which is already utilized by these [1],[2] paper.

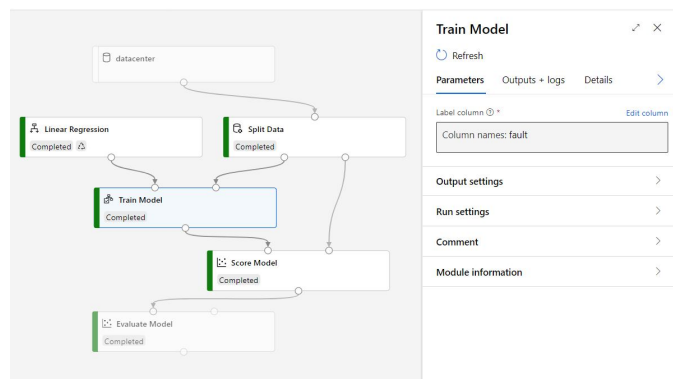
### 3. PERFORMING THE LEARNING REGRESSION

First uploading the CVS file to the dataset of Microsoft Azure Machine Learning Studio. Moving forward to the technical aspect to Feature out which column we will work on or not. The more incorrect data generate the typical errors in the training model phase. As taking the dataset and grab it to the Microsoft Azure Machine Learning Studio canvas. Then it can see that applied Spilt Data Model to figure out the spilled data into the Test and Train samples. The most crucial thing it can be adjust its ratio of 0.5 to 0.7. But it can get this thing is most important for checking the sampling amount if you have insufficient data records or have a larger dataset.



**Figure 1:** Experiment Workflow on Microsoft Machine Learning Studio set the Split Data ratio.

Mostly, having around a total of eleven thousand documents in CSV format. The column is serial, product, and fault.



**Figure 2:** Experiment Workflow on Microsoft Machine Learning Studio set the Label.

The above chart shows that the most common fault in the data center is the Hard Drive of systems; the seconds most common issue is the Switch failure, the third one is the network problems, next is the ram issue, and the last one is the SSD issue. These are the widespread issue faced by the data center in Karachi, Pakistan.

#### 4. CONCLUSION

The above procedure shows that the datacenter in Karachi, Pakistan is having five significant issues when it's come to the High Availability of Datacenter. If we will provide a way to check, it's a failure before the mainstream of datacenter.

#### 5. FUTURE WORK

We will highly encourage you to be part of our research. For this, you need to participate in it. Our next goal is to enhance the feature by using the SNMP logs of any Karachi data center. For this first, we require the dataset as once we have

via the Premier Systems (Pvt) Ltd. Although, we could find the way and then build this as well.

#### ACKNOWLEDGEMENT

We want to thank first Premier Systems (Pvt.) Ltd. to support this research and give its organization big data. Special thanks to Mr. Sahal Qasim (email: sahal.qasim@premier.com.pk) to cooperate with us to deliver the data. Moreover, this research outcome will also help Premier Systems (Pvt) Ltd. to figure out which are the gaps is in the physical on-premises data center in Karachi, Pakistan. I would, thanks to my supervisor, for encouraging me to do the research. Especially thanks to my family and colleagues; without their support, this would not be achievable.

#### REFERENCES

1. Botchkarev, Alexei, Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio (May 12, 2018). Available at SSRN: <https://ssrn.com/abstract=3177507> or <http://dx.doi.org/10.2139/ssrn.3177507>
2. Botchkarev, A. (2018). Revision 2 Integrated tool for rapid assessment of multi-type regression machine learning models. An experiment in Microsoft Azure Machine Learning Studio. Azure AI Gallery. <https://gallery.azure.ai/Experiment/Revision-2-Integrate-d-tool-for-rapidassessment-of-multi-type-regression-machine-learning-models>
3. Shuigeng Zhou, Jihong Guan, Evaluation and Construction of Training Corporuses for Text Classification: A Preliminary Study Natural Language Processing and Information Systems, 2002, Volume 2553
4. R. Dennis Cook (1977) Detection of Influential Observation in Linear Regression, *Technometrics*, 19:1, 15-18, DOI: 10.1080/00401706.1977.10489493
5. Evaluate Model. (2018). Azure Machine Learning Studio. [https://docs.microsoft.com/enus/azure/machine-learning/studio-module-reference/evaluate-model#bkmk\\_regression](https://docs.microsoft.com/enus/azure/machine-learning/studio-module-reference/evaluate-model#bkmk_regression)
6. Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669-679.