Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013

Dr. Germán H. Alférez, *Universidad de Montemorelos,* Erón Zebadúa, *Inter-Oceanic Mexican Union Conference,* and Enoc Cruz, *Universidad Linda Vista*

Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos

Abstract—Removing individuals from membership in the Seventh-day Adventist Church is the ultimate discipline that the church can administer. Our contribution is to present how we have applied state-of-the-art data science techniques to identify the segments of individuals who have been baptized from 2005 to 2013 and also been removed from membership in the same period of time at the Inter-Oceanic Mexican Union Conference. The dataset that was analyzed is composed of 14,388 records of members who have been removed. The results can guide further church decisions to prevent membership lost, specially among youth and among people who are baptized after evangelistic campaigns. Our data-science approach could be easily extrapolated to other divisions and conferences.

Index Terms—data science, Seventh-day Adventist Church, machine learning

✦

1 INTRODUCTION

No organization can long survive without retaining and finding customers. In our church context, our members are our customers. The process of retaining and finding new members begins by learning as much as we can about current members and groups of members.

In this work, we identify the types of members who are most likely to leave the church as well as sets of variables that may be used to find those members. Ranges of values on those sets of variables define church segments, which in turn can be used to identify members who are at risk of leaving the church.

A church segment is a group of members that is different from other groups of members in ways that matter to church administrators and pastors. For church retention, a person's segment membership can be among the most useful explanatory variables. The process of identifying church members is called segmentation.

Our contribution is to show how we have applied data science techniques to identify segments of individuals who have been baptized from 2005 to 2013 and also been removed from membership in the same period of time

- *Dr. Germán H. Alférez, School of Engineering and Technology, Universidad de Montemorelos, Montemorelos, Mexico*
  E-mail: harveyalferez@um.edu.mx, Website: www.harveyalferez.com.
- *Erón Zebadúa, Inter-Oceanic Mexican Union Conference of the Seventh-day Adventist Church, Puebla, Mexico*
  E-mail: eron.zebadua@gmail.com.
- *Enoc Cruz, Universidad Linda Vista, Pueblo Nuevo Solistahuacán, Mexico*
  E-mail: ecruznajera@ulv.edu.mx.

at the Inter-Oceanic Mexican Union Conference of the Seventh-day Adventist (SDA) Church. Data science can be defined as the study of the generalizable extraction of knowledge from data [1]. As far as we know, this is the first study in which machine-learning techniques are used in the context of data science to understand the segments of members who have been removed from membership in the SDA Church. The results are key to recognize segments of individuals who may be at risk of leaving the church in the future. The underlying machine-learning techniques that were used in the experiments could be easily applied by other divisions and conferences.

This paper is organized as follows. The second section presents why data science was chosen to understand the segments of members who leave the church at the Inter-Oceanic Mexican Union Conference. The third section presents the description of the dataset that was analyzed in this study. The fourth section presents the results of our experiments. Finally, the fourth section presents the conclusions.

2 WHY DATA SCIENCE?

In this study, we use cluster analysis to identify the segments of individuals who leave the church at the Inter-Oceanic Mexican Union Conference. Cluster analysis involves finding groups in data. Cluster analysis takes many member-related variables and uses them to represent differences between members. When two

members have similar values for variables like age and religious background, they are seen as being similar. When two members have very different values on these variables, they are seen as dissimilar. Cluster analysis looks at the differences among members–their distances from one another–to identify groups of members.

It is important to clarify the following aspects related to the role of clustering and machine learning in data science:

- Within the machine learning literature, cluster analysis is referred to as unsupervised learning to distinguish it from classification, which is supervised learning, guided by known, coded values of a response variable or class. Thinking broadly about machine learning, we see it as a subfield of artificial intelligence [2], [3].
- We use the term "machine learning" to refer to the methods or algorithms that we use as an alternative to traditional statistical methods. When we apply these methods in the analysis of data, we use the term "data mining." Rajaraman and Ullman [4] describe data mining as the "discovery of models for data."
- With traditional statistics, we define the model specification prior to working with the data. With machine learning, the model specification is defined by applying algorithms to the data (i.e., data-adaptive). With machine learning, few assumptions are made about the underlying distributions of the data.
- Data science is the new statistics, a blending of modeling techniques, information technology, and business savvy [5]. Machine learning is a key component of data science, which lays at the intersection of programming skills and math & statistics knowledge.

3 DESCRIPTION OF THE DATASET

The dataset that was used in the experiments is composed of data from 14,388 members who were baptized and removed from membership from 2005 to 2013 at the Inter-Oceanic Mexican Union Conference. This Union Conference was established in 2001. It covers the following Mexican territory: Guerrero, Hidalgo, Morelos, Oaxaca, Puebla, Tlaxcala, the southeast portion of Tabasco up to the river of Samaria and Mexcalapa, and Veracruz (except the municipality of Xochitlan); comprising the Isthmus, North Veracruz, Oaxaca, Olmeca, South Pacific, South Veracruz, and Southeast Veracruz Mexican Conferences; and the Alpine, Central Veracruz, Chontalpa, and Los Tuxtlas Missions. This Union Conference has 1,421 churches and 197,890 members [6].

The dataset that was used in the experiments was taken from a membership database created in-house by the Inter-Oceanic Mexican Union Conference staff in 2004. This database was updated by the conferences' and missions' secretaries in the territory of this union conference until the year 2013. Afterwards, this database

was replaced by the Adventist Church Management System (ACMS)[1]. Although the in-house system was created in 2004, we used the records from the year 2005 to 2013 for the experiments since results in 2004 were incomplete.

The dataset that was used for the experiments has the following fields: religious background before baptism (Seventh-day Adventist, Catholic, Pentecostal, and none), ways in which the individual was converted (by an evangelistic campaign, by a Bible course, or by a friend), date when the member was baptized, and date when the member was removed from membership.

4 RESULTS

This section presents the results of 3 experiments. The experiments were executed on Weka[2]. The k-means algorithm[3] was applied on the data. This algorithm finds k clusters (church segments in our case) for a given dataset. The number of clusters k is user defined [7]. Therefore, in the experiments we searched from solutions with between 3 and 6 clusters. We used the number of clusters that made the most sense. In the experiments the within cluster Sum of Squared Errors (SSE) was 40.55 in the first experiment, 402.80 in the second experiment, and 134.83 in the third experiment[4].

- FIRST EXPERIMENT: 57.21% of members who are removed from membership are young (34.15% teenagers and 23.06% youth) and last around 3 years at church (see Table 1). In fact, independently from the age group, members who leave the church last in average 3.26 years.

Table 1
Percentage of Church Abandonment by Age Group

Age Range	Percentage of Abandonment	Amount of Members who Abandon	Duration in Years as Church Member
12-16	34.15	4,914	3.35
19-22	23.06	3,318	3.13
28-32	16.68	2,400	3.10
39-42	12.53	1,803	3.08
53-56	8.85	1,273	3.20
70-74	4.73	680	3.68

- SECOND EXPERIMENT: Table 2 shows the percentage of the retention of youth (from 15 to 33 years old) by religious background before baptism. The percentage of retention of youth with a SDA background is very similar to the segment of youth without any religious background. In other words, the chance of

1. ACMS: https://www.acmsnet.org/
2. Weka: http://www.cs.waikato.ac.nz/ml/weka/
3. Although the k-means algorithm can be executed on Excel by means of particular plugins, there are many more clustering algorithms than k-means. Weka and other machine learning tools are more appropriate and efficient than a spreadsheet software to run them.
4. SSE measures the variation within a cluster. SSE would be equal to 0 if all samples within a cluster are identical.

a young individual with a SDA background (e.g. who comes from a SDA family) to stay at church is very similar to the chance of a young individual without any religious background. Also, it is interesting to see that, one hand, the segment of youth with a Catholic background has the highest chance to stay at church. On the other hand, the segment of youth with a Pentecostal background has the lowest chance to stay at church.

Table 2
Percentage of Youth Retention by Religious Background

Religious Background	Amount of Baptized Young Members	Among of Removed Young Members	Percentage of Retention of Young Members
Pentecostal	1,191	441	62.97
Seventh-day Adventist	18,677	4,246	77.26
None	26,409	5,359	79.70
Catholic	6,860	1,055	84.62

- THIRD EXPERIMENT: People who take formal Bible courses before baptism or who are invited by a friend to come to church are less likely to leave the church than people who enter the church after an evangelistic campaign (see Table 3). It is important to mention that people who are baptized in an evangelistic campaign do not necessarily have to take a Bible course.

Table 3
Percentage of Retention by the Way of Entering the Church

Way of Entering the SDA Church	Amount of Baptized Members	Amount of Removed Members	Percentage of Retention
Evangelistic Campaign	32,553	9,488	70.85
Bible Course	2,507	362	85.56
By a Friend	39,639	4,538	88.55

5 CONCLUSIONS

This paper presents how we have used data science to try to understand individuals who have been removed from membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013. Specifically, we have found the following results: 1) most individuals who are removed from membership are young and last around 3 years at church; 2) the percentage of retention of youth (15 to 33 years old) with a SDA background is very similar to the segment of youth without any religious background. Also, the segment of youth with a Catholic background has the highest chance to stay at church and the segment of youth with a Pentecostal background has the lowest chance to stay at church; and 3) people who enter the church after an evangelistic campaign tend to leave the church in a higher percentage than people who take Bible courses or who are invited by friends.

If we are successful in identifying differences across church segments, then church administrators and pastors will have an easier time understanding the meaning of segments. And if managers and pastors understand the meaning of segments (how segments are defined, how they differ from one another), then they will be more likely to use the results of segmentation to guide church decisions to prevent membership loss.

ACKNOWLEDGMENT

REFERENCES

[1] V. Dhar, "Data science and prediction," *Commun. ACM*, vol. 56, no. 12, pp. 64–73, 2013.
[2] G. Luger, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 6th ed., Boston, Ed. Addison-Wesley, 2008.
[3] S. N. Russel, *Artificial Intelligence: A Modern Approach*, N. Upper Saddle River, Ed. Prentice Hall, 2009.
[4] A. Rajaraman and J. Ullman, *Mining of Massive Datasets*, U. Cambridge, Ed. Cambridge University Press, 2012.
[5] T. Miller, *Marketing Data Science*, Pearson, Ed. Old Tappan, 2015.
[6] Office of Archives, Statistics, and Research, Seventh-day Adventist Church. (2016) Inter-oceanic mexican union conference. [Online]. Available: http://www.adventistyearbook.org/ViewAdmField.aspx?AdmFieldID=IOMU
[7] P. Harrington, *Machine Learning in Action*. Shelter Island. N. Y., 2012.

Germán H. Alférez is a professor at the School of Engineering and Technology, Universidad de Montemorelos, Mexico. He holds a Ph.D. in Computer Science from Universidad Politécnica de Valéncia (Spain), a MSc in Information and Communication Technology from Assumption University (Thailand), and a BSc in Computer Science Engineering from Universidad EAFIT (Colombia). His research interests include Services Computing, Model-Driven Engineering, Models at Runtime, Autonomic Computing, Data Science, and Dynamic Software Product Lines. He has contributed to publications in journals, book chapters, and international conferences, such as SPLC, ECOWS, MODELS, ICWS, IUPESM WCMPBME, HIMS, ICAI, and SERP. He has worked in universities, IT companies, and research groups of four continents (America, Asia, Australia, and Europe). He leads the Global Software Lab at the School of Engineering and Technology, Universidad de Montemorelos.

Erón Zebadúa holds a BSc in Computer Science from Universidad de Montemorelos, Mexico. He works at the Department of Communication at the Inter-Oceanic Mexican Union Conference. He has served at the Seventh-day Adventist Church for 17 years. During this time he has developed a membership management software, a treasury software for local churches, and academic software for schools in this union conference. He is a passionate programmer, a Linux enthusiast, a photographer, and a video and audio editor. Currently he is studying a MSc in Computer Science at Universidad de Montemorelos, Mexico.

Enoc Cruz is a lecturer of databases, programming, and software engineering at the School of Engineering at Universidad Linda Vista, Chiapas, Mexico. Also, he is the coordinator of the Research Center of Information Technologies at this school. He holds a MBA from Universidad de Montemorelos and a Microsoft Certified Database Administrator (MCDBA). Currently he is studying a MSc in Computer Science at Universidad de Montemorelos, Mexico.