Utilizing Video Views to Create Recommendations for Marketing Funnels

Patrick W. Delaney

Abstract**—** Clustering is a useful tool to make serendipitous or at least random recommendations for marketing segmentation. In this paper we propose a methodology to translate video views into audience segmentations for digital marketers, knowledge workers or others whose job it is to make sense of industrial markets. Using a method known as Principal Component Analysis combined with K-Means clustering and an abstract control systems method to improve results, it was observed that useful segments can be built based upon an audience's viewing behavior. Further study and practical application is required to expand this effort.

Index Terms**—**Machine Learning, Dimensionality reduction, Machine learning algorithms, Clustering methods, Clustering algorithms, Classification algorithms, Knowledge management, Marketing management, Information filtering, Recommender systems, Videoconferences

I. Introduction

DIGITAL marketers, data scientists and generally those tasked with characterizing entire industries and sectors are working in a world of ever-expanding and overwhelming data and choices. Mapping out audience or industry stakeholder interests is increasingly complex, and there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload.

Recommender systems, and in particular those which employ clustering algorithms help solve this problem by searching through large volumes of dynamically generated information to provide users with personalized content and services [1]. Likewise, Clustering is a highly popular and widely used tool for identifying or constructing databased market segments [2]. Fig 1. demonstrates a standard business funnel model for turning prospects and leads into buyers, clickthroughs or other actions. Typically in any business environment, the top of the funnel includes a larger volume of professional contacts, and contains a high number of potential customers or users. Moving down the funnel, there may be various abstract stages of increasing interest, ultimately leading to an action.

Recommender systems can extract information from the process of converting a possible user from, "higher in the funnel," to, "lower in the funnel," by looking at cues from

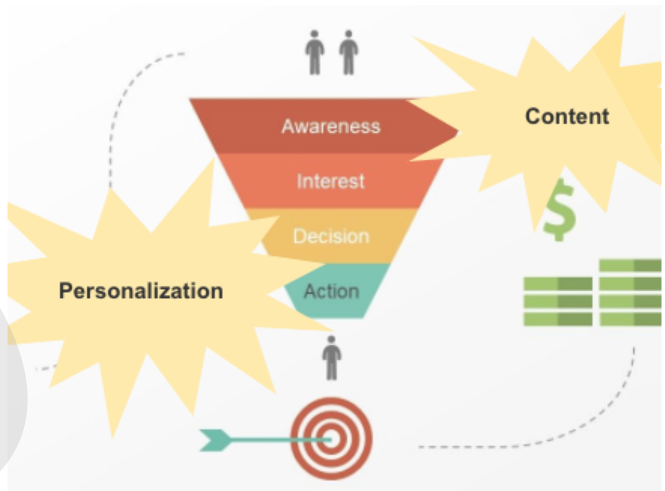

Fig 1. Standard business funnel models for turning prospects and leads into buyers, clickthroughs or other actions.

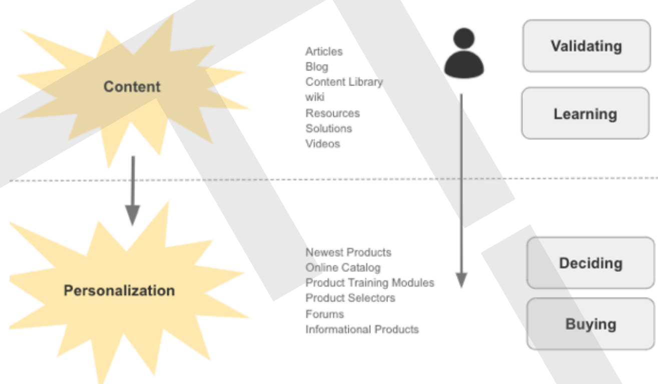

Fig 2. A fundamental part of characterizing an industry segment is being able to bridge the gap from content ingestion, or more abstract, "learning" activities to, "deciding," activities such as buying or selecting.

Manuscript not submitted. This work is self-supported.
More work and contact information for Patrick W. Delaney may be found at www.patdel.com

how content is discovered and consumed. The job of a marketer or digital content professional is ultimately to find patterns in how content is being consumed, "higher in the funnel," and increase the efficiency of converting users to, "lower levels of the funnel." How, when and to what degree users engage with content are all points of information that can be used to help bridge the gap from diffused, convoluted information to informed decisions as shown in Fig 2. Moreover, using the appropriate clustering algorithm, digital marketers can be fed, "random / serendipitous" recommendations which they may not have realized purely through human analysis - which we discuss further in Section III A.

A. The New Role of Video Content in Creating Recommendations

Even before the 2020 pandemic, the late 2010's ushered in an age of, "near-ubiquitous online video." The ability to quickly and easily stream and share video content began to take prominence on the web. Services such as Twitch.tv, as well as live streaming video offerings from Facebook, YouTube, Tik-Tok and Twitter have brought easy video creation and ingestion into the public consciousness. "The Creator," or, "The Teacher," has taken center stage in the early 2020's as the pandemic has pressed people into social distancing situations away from schools, offices and public venues.

Interestingly, research has already been done on the role that online video can play from a psychological perspective to viewers. Two key interesting findings about online video have been observed methodologically:

1. Video watching helps in forming positive emotional response by connecting individuals with brands, rather than faceless names, numbers, parts [3].
2. Learning outcomes and satisfaction improves with video lesson and lecture watching [4].
3. Emotional qualities of brand value play an important role in Business to Business (B2B) marketing. Forming an emotional connection with buyers is important in B2B [5].
4. Emotional connection with a brand facilitates the progression from goods and services value to loyalty, better profitability, referrals, and the ability to cross-sell other products and services [5].

Essentially, studies on video viewing for online learning have shown that individuals who watched an above-threshold of video minutes improved how much they learned about a topic, regardless of where they started out in their knowledge level. In addition to learning and absorption, emotional satisfaction in brands and companies improved upon watching videos which has been shown to improve loyalty, profitability and cross-selling.

A key assumption we make drawing off of these combinations of studies is that higher engagement in video viewing, among videos that are designed to teach about a new topic, product or service shows a greater level of learning and positive emotions and therefore engagement with an industry offering, service offering or product. In essence, we can say that viewing the right videos can be considered to be, "similar to" future buying of products from a given brand, as it leads to higher loyalty and purchasing.

In other words, video viewing behavior can be considered to be, "similar" to purchasing or deciding, from a psychological standpoint. Translating this psychological assumption into a mathematical framework, we use some form of cluster analysis. Cluster analysis is the mathematical task of grouping observations in such a way that items which are close to each other in value, distance, measurement or by some pre-defined method are in the same group.

Using a psychological approach, we are making assumption drawing from the previous studies mentioned that a, "product purchase" is similar to a "video view," because a user has to invest time into viewing a video on a particular topic.

In using cluster analysis, we are making the philosophical assumption that sets of observations about online video views found to be similar to one another in a mathematical sense will also be similar to one another in a psychological sense. It is important to draw the distinction between our psychological assumption and our philosophical assumption. Cluster analysis and usage of clustering algorithms to describe psychology or human behavior will always yield a result. It is important to be clear that this result is only a model. The model can be, "improved" by selecting and refining algorithms for efficiency, creating more well-defined segments and, "cleaning up," the groups to make sure they are more distinct. However, the idea of using cluster analysis to transform, "top of the funnel," marketing and further, predict activities further at the, "bottom of the funnel," relies not only on appropriate algorithmic modeling using the appropriate clustering techniques to model video viewing behavior, it also relies on the assumption that video viewing behavior can be considered to be, "similar" to purchasing or deciding activities.

B. About Recommender Systems and Clustering

"Clustering" is a group of different mathematical methods which turns observations into clusters. In a sense, Clustering is a family of algorithm types. In contrast to this, "Recommender systems," are more of a combination of ways to create recommendations, which utilizes algorithms which may include but are not limited to clustering. Recommender systems can be as complicated or as simple as they ultimately need to be and can be designed in different ways. Recommender systems are nothing, "magical," in fact they have been used since the dawn of time. A pre-historic human telling their friend, "I recommend only flat gray rocks with no moss growing on them to make the best arrowheads," is an example of constructing a recommender system which one can carry forward in the mind.

Today, recommender systems are widely used on the web. The most typical example is of course the E-Commerce based recommender system which makes recommendations along the lines of, "customers who bought X also bought Y."

However, other things we use on a daily basis may also be considered recommender systems:

1. Search: Search is essentially a ranked choice way of recommending something from a massive group. Search typically shows the top recommendation at the highest level, followed by the second and so on. No one is obviously expected to sift through every single possible result, but rather look at the, “top,” of the list and decide from there.
2. Security Risk Identification: In massive, highly critical systems being monitored by specialists, such as online banking systems, massive e-commerce operations or even networks of cameras, gates and locks which spread out over thousands of locations - there may be huge numbers of false positive and true positives occurring on a per-minute if not per-second basis. There is simply not enough people in the world to manually check every single potential security thread in existence, so much of large system security today is achieved through adaptive recommender systems.
3. Medical or Healthcare Pre-Diagnosis: Particularly during the current pandemic at the time of authoring this article, health systems are often pushed to the brink of not being able to efficiently process every single possible condition for every possible patient in existence. Over the past decade or so, many digital health initiatives have sprung up which utilize a range of either clinically verified or non-clinically verified ways of pushing human behavior in ways that improve health outcomes, whether it be by recommending caretaker visits or highlighting risk factors for certain diseases.

Recommender systems are not the, “end-all-be-all,” of all things digital. Effective digital product design and deployment involves a combination of psychology, user experience discipline, software skill, effective project management, and other forms of know-how. However, if properly designed they are certainly an important mathematical and user interface practice which can help reduce an arena of complicated, chaos into one that is more, “human,” and which can help individuals and teams make more informed decisions.

This paper will go through an example of how a clustering technique, combined with a control systems inspired decision feedback loop, can create recommendations for actions that digital marketers can take to improve conversion efficiency in a business funnel. It is important to note that while this recommender system approach may be fairly generalizable, but the clustering algorithm used must be selected and applied with caution.

As mentioned in Dolnicar (2002), “The application of cluster analytic procedures for the purpose of data-driven segmentation studies should become much more careful in the setting of parameters in order to substantially improve the quality of clustering outcome and reduce the proportion of “random results” which are interpreted in detail and misunderstood as best representation of the data in reduced space,” [2].

II. Preparing the Data for Clustering

A. Overview of Clustering Audiences via Viewership Interest

Typically recommender systems are thought of as systems that recommend a, “result” from a store, large database, or huge list to an individual. As demonstrated in Fig 3., recommendations could also be given based upon not just a store of, “products” or “videos” but also based upon behaviors of a group. In other words, the actions of a large number of individuals are sort of like the, “products on the shelves,” and a mathematical analysis of these large number of actions can produce a recommendation to the effect of, “here’s how to group these folks together,” or, “here is a pattern you may not have considered. These recommendations can be combined with human expertise to optimize results.

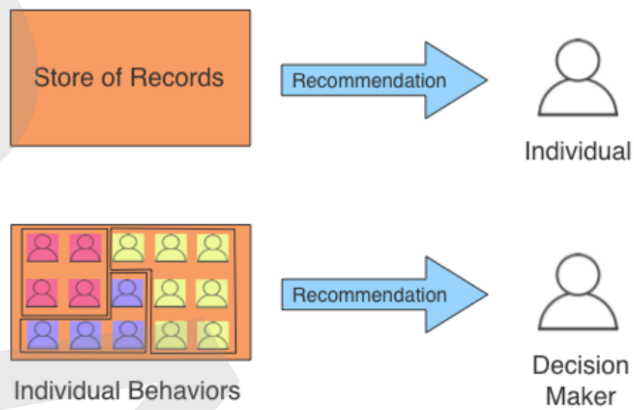

Fig 3. Recommendations could also be given based upon not just a store of, “products” or “videos” but also based upon behaviors of a group.

B. Information Collection

The data generated for this paper was from a piece of software the author created and owns called, “Confrnz.” Confrnz is basically a live-streaming and video viewing software, which includes a variety of different workshops and conference style talks given on technical and industry trends within the, “Internet of Things,” market sector. This market sector is large and varied and videos were shown across multiple industry segments and a wide range of topics over 9 months to a subscriber base of around 1000 viewers.

Viewers are able to log in and view videos directly on the site as shown in a manner similar to what is seen in Fig 4., with their clicks and actions tied to their independent identification. Users could click on individual videos by topic, enter video, “pages,” and choose to either remain and watch the videos or navigate away. Videos were varied over time, and individual viewership of the videos was tracked and logged in a manner discussed later in this paper.

After a user logs into the software, their viewing history is tied to them by email, and a unique user I.D., as shown in

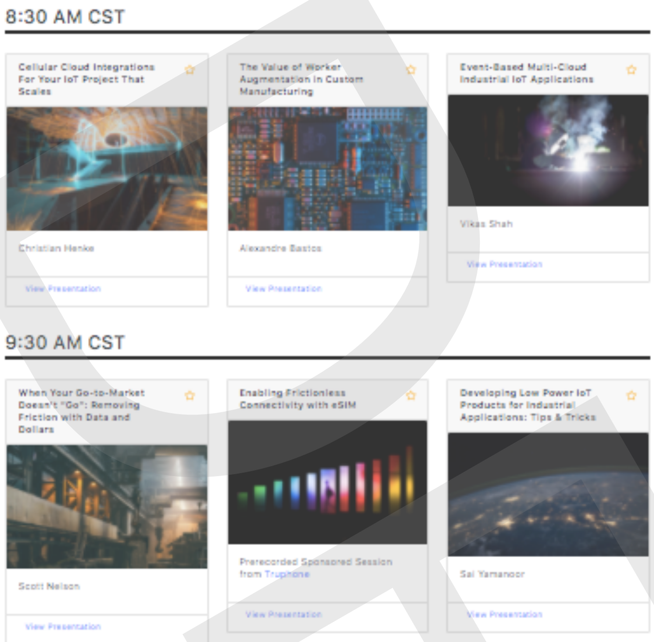

Fig 4. Layout of individual videos menu within the software used to collect viewership information for this paper. This shows a typical way that different options were displayed to viewers. Upon clicking a selection from the menu, a user navigates to the video page.

Fig 5. By logging all user video views, a large summary matrix of all videos vs. all users, with the values in the matrix corresponding to view times can be built which can be used for further cluster analysis.

C. Implicit User Feedback vs. Explicit User Feedback

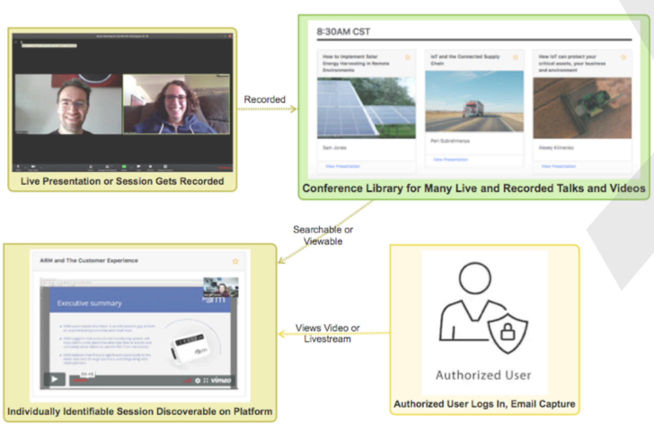

Fig 5. After a user logs into the software, their viewing history is tied to them by email, and a unique user I.D., the user is then able to view any videos they choose from a library of pre-recorded talks and workshops.

Reducing complexity for marketers and decision makers can be done by looking at user behavior and grouping users based upon their actions. This might be termed something along the lines of, "behavioral grouping."

Information can be tracked from users in a couple different ways:

1. Explicit user feedback, such as rating scales, likes, saves or anything that requires a decision beyond navigation provides users with a mechanism to actively, "push buttons" to show their items.
2. Implicit user feedback is generated by observations of the user behavior itself, for example, whether someone clicked on something at all, or time spent on a page.

Our experimental setup consisted exclusively of implicit user feedback, video views, although explicit user feedback could be modeled into a cluster analysis as well. Table I shows the implicit feedback gathered with our software. The data above represents, "click" data which is implicitly collected. When a user lands on a particular page, then that user's click is measured at, "Timestamp," for that particular page. When they click off the page (but stay on the platform), that click is measured as, "Exit Time," whereas

TABLE I
IMPLICIT USER FEEDBACK GATHERED

Datapoint	Description
EntryID	Number designating row in entire data table.
UserID	User ID number for a particular user
Session	Session unique name, identifies the video
Timestamp	Entry time, point of entry on webpage, measured as soon as page starts load
ExitTime	Exit time, point of exiting webpage upon clicking away from webpage

Table 1 shows raw datapoint rows captured for each session visit and exit.

complete exits from the page are not registered at all. This is in line with how Google measures time on page or, "average session duration." Overall, some kind of fundamental definition of a, "Video View" can be defined mathematically, with pre-loading intervals and dwell times which can be tweaked to translate our above mentioned psychological assumption that, "a video view indicates interest," into a mathematical framework, as shown in Fig 6. It is critical to break down how user inputs and time on page concretely links to user interest. While it may be tempting to think of a view as a binary occurrence, e.g. either a, "view or not view," data captured shows an exponential distribution as shown in Fig 7. and Eq. (1) for the number of users on a given page and the amount of time spent on that given page. This exponential distribution makes intuitive sense as video viewing time is a limited resource.

Eq. (1)

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Once a concrete definition is decided upon and with time on page collected for every individual user, it is possible to build a large matrix, as demonstrated in Table II, a "User

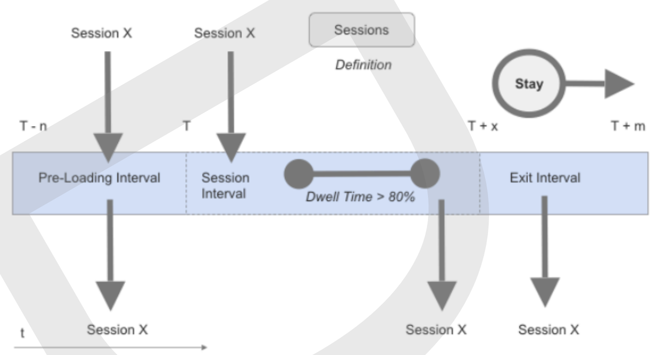

Fig 6. "Video Views" can be defined, with pre-loading intervals and dwell times which can be tweaked to better define interest and purchasing intention in a mathematical framework.

Time on Page Matrix," capturing whether or not and for how long every individual user has, "watched," any given video within a conference. This dataset may be either transformed into binary, "Watched" or "Did Not Watch," values or simply analyzed closer to "Raw Duration," format, which, given Eq 1. derived from Fig 7., seems that it may give a more accurate segmentation end result.

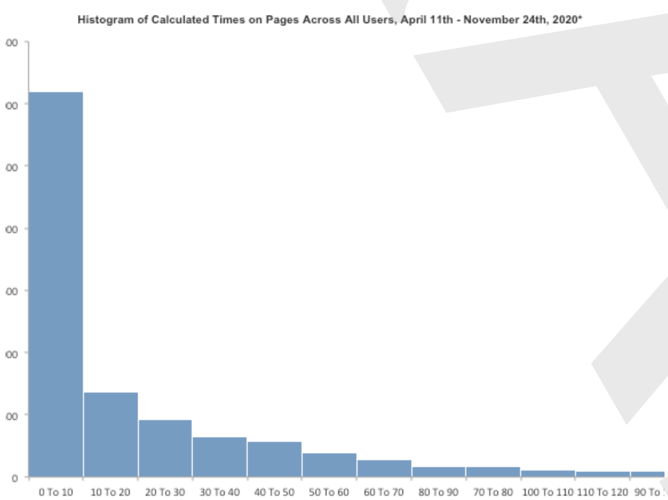

Fig 7. Histogram of Calculated Times on Pages Across All Users, April 11th - November 24th, 2020. Note: Time on Page Durations of less than 1 minute and greater than 120 minutes were eliminated for this chart. Durations of more than 120 minutes constituted less than 1% of all views.

TABLE II
SAMPLE USER TIME ON PAGE MATRIX

	Session1	...	SessionN
User1	2 minutes	30 minutes	0 minutes
User2	2 minutes	2 minutes	2 minutes
...	...	...	...
UserN	2 minutes	15 minutes	2 minutes

Table II shows raw datapoint rows captured for each session visit and exit.

D. Ensuring Sufficient Data

One of the key questions for any data science project is, "how much data do you need to make an effective prediction?" The short answer is invariably, "it depends." It depends upon the data problem you are looking to solve. Some examples may include:

1. Image classification, depending upon the complexity, may require thousands or more images to train a, "classifier." Of course this could highly depend upon what exactly you are trying to identify in the images, the number of pixels, the range of colors and brightness, and how accurate the model needs to be.
2. For regression, type problems, a general rule of thumb is that you should have at least ten times more observations (data points) than you should features (or variables in the equation you create). So for example if you create a line with the formula $y = mx+b$, given that there is one (1) variable, you should probably have around 10 observations. Obviously this depends upon the, "tightness," of the data and what the problem you are trying to solve really calls for as well.
3. For clustering type problems, the number of samples needed to build an effective cluster is actually dependent upon the number of clusters that will be built. The more clusters or groups that are needed, the more data that is needed, in general. Different types of clustering algorithms can be chosen based upon how much data is available.

Basically, the amount of data needed is largely a philosophical problem - there is no, "Golden Rule," of data needed to make an effective prediction or quality recommendation.

Since we are talking about clustering problems for market segmentation specifically, it is important to note that, clustering techniques will, "always create a result," regardless of whether that result makes sense or not. What gathering more behavioral and better quality data does is reduce the proportion of, “random results,” which can misrepresent what is happening in reality.

That being said, with the Confrnz system that was built and tested starting in April, 2020 through November, 2020, significant implicit data capture capability and the ability to increase the amount of data intake reliably over time was demonstrated.

By running regular small conference talks and workshops with about 10 to 15 speakers per event, with events happening each month, we were able to collect around 33 to 35 lines of data per day on average throughout the time period in question. The lines of data included both entry times and exit times tied to userID as shown in Table I. This works out to about 85 lines of data per speaker on our software platform at this time.

As shown in Fig 8., Over the course of 160 days, starting on 4/23/2020 and ending on 11/24/2020, the platform we built was able to collect 10,403 lines of data. The large spike at the beginning was due to a large amount of data collected during a highly attended event.

Since our data collection is ongoing, and based upon regular roll-out of video content, there is an opportunity to consistently develop new results, and re-apply the clustering algorithm as the world changes. Decaying time-series analysis, either linear, polynomial or exponential may be applied as necessary, putting a lower, “weight,” to video views which happened in the past. We will discuss a control-systems approach to adapting the clustering algorithm later in this paper.

III. Performing the Cluster Analysis

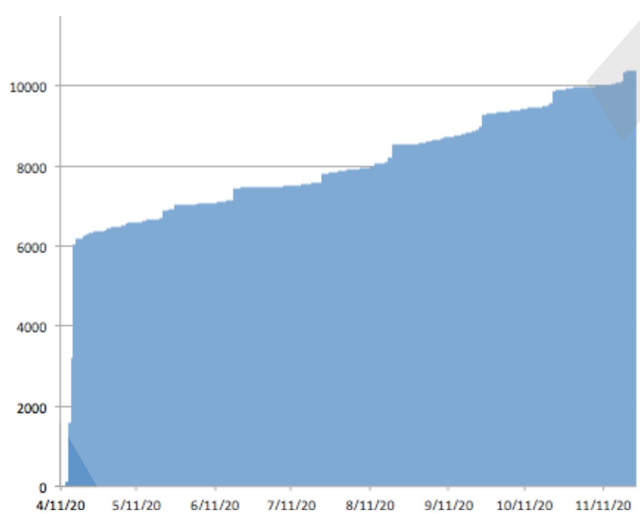

Fig 8. Lines of data collected between 23 April 2020 and 24 November 2020. The large spikes represents a larger event with many videos, the smaller spikes represent a smaller event with fewer videos.

A. Collaborative Filtering vs. Content-Based Filtering

Besides the mathematical methods involved in cluster analysis, there are a couple different philosophies on how this clustering or classification can be performed for recommendations [6].

1. Collaborative Filtering recommends items by identifying other users with similar taste; it uses their opinion to recommend items to the active user. Collaborative recommender systems have been implemented in different application areas.
2. Content-Based Filtering techniques match content resources to user characteristics. Content-based filtering techniques normally base their predictions on user’s information, and they ignore contributions from other users as with the case of collaborative techniques.

Let x be a zero-mean random variable. Suppose we want the direction w such that the projection of x along this direction has maximum variance:

$$\max_{w}(w'x) \quad st. \quad w = 1. \tag{1}$$

We have

$$(w'x) = w'xx'w = w'\Sigma w. \tag{2}$$

The Lagrangian is

$$L = w'\Sigma w + \lambda(w'w - 1). \tag{3}$$

The stationary condition is

$$\frac{\partial L}{\partial w} = 2\Sigma w - 2\lambda w = 0, \Sigma w = \lambda w. \tag{4}$$

Thus w is an eigenvector of Σ. Since

$$w'\Sigma w = w'(\lambda w) = \lambda, \tag{5}$$

the direction with maximum variance is the largest eigenvector. This procedure can be iterated to get the second largest variance projection (orthogonal to the first one), and so on. For a set of data points, we use the ML estimate of the covariance matrix.

Fig 9. Mathematical Shorthand derivation of principal component for a data set of click data or, “random variable.” With click or view data, the random variable x would likely not have a zero mean. We can find the dimensions of the PCA by using the covariance matrix of x.

Generally, “Collaborative Filtering,” requires more data, and requires a lot of inter-relationships between the data, but if done properly and with sufficient data it can be used to create, “serendipitous recommendations,” which may not have been identified by digital marketers through human analysis. Collaborative Filtering looks at the user feedback, either implicit or explicit, and finds patterns within that data. In contrast, content-based filtering, “tags” data based upon presumed interests and fields, or asks users to fill out surveys, and builds recommendations based upon this feedback. Collaborative Filtering, in a sense, suggests tags for humans to evaluate, whereas content-based filtering uses human-build tags as inputs, which are grouped by computers. With the relatively large amount of data we were able to gather, and with the goal of creating serendipitous recommendations

TABLE III
COMPARISONS OF SIMILARITY METRICS

Name	Shorthand	Example Application	Rank Items
Cosine of Vector Angle	$\cos\varphi = \frac{\vec{A}\cdot\vec{B}}{\left\|[\vec{A},\vec{B}]\right\|}$	Amazon 1.0 Product Recommendations	Rank items according to top items purchased by similar customers, by distance calculation.
Naive Bayes Classification	$P(X,Y\|Z) = P(X\|Y,Z)P(Y\|Z) = P(X\|Z)P(Y\|Z)$	Predicting Whether Cloudy or Sunny Tomorrow	Classifying units into buckets A, B, C, etc. by presumed probability.
Linear Prediction Coefficient	$y[n] = \sum_{p=0}^{P} a_p x[n-p] + \sum_{q=0}^{Q} b_q y[n-q]$	Speech Filtering, Classifying Tumor Danger by Size	Creating a boundary condition and classifying on either side.
Decision Trees	$E(S) = \sum_{i=1}^{c} -p_i\left(log_2\right) p_i$	Picking a real world thing from a list, such as clothes to wear based upon today's weather.	Minimize, "pain" based upon weighted factors.
Hamming Distance	$H(x,A) = \min_{y\in A} H(x,y)$	Telecommunications, flipped bits.	Minimum number of substitutions required is, "best."

Table III shows different types of similarity metrics and clustering models.

our analysis uses collaborative filtering. There are of course, disadvantages to collaborative filtering, which we will discuss in Section V.

B. Selecting Similarity Metric

Since our software platform, Confrnz uses Collaborative Filtering, the next step is to decide which type of Collaborative Filtering will be employed, or rather, which type of algorithm may be the most appropriate for the situation that will result in the best cluster analysis or market segmentation recommendations given to the marketing teams.

We should take due care when applying our algorithm to the data we have on hand, depending upon the size of the data set, number of variables, associated data format and so on [2]. It is important to choose the right type of algorithm. We start of with the, traditional collaborative filtering algorithm, which is used by large, established companies for product recommendations when comparing customers to one another. Caution must be used to ensure the analysis does not suffer from a sparse dataset, but for our application, though it may be computationally intensive, it can happen on a server, "in the background," as an instant or browser-reactive—based result is not necessary.

From Table III. we select, "Cosine of Vector Angle," or "Euclidian" distancing since this is the most, "traditional" type of similarity measurement that is used in product recommendations in terms of describing customer similarity to one another [1]. We are making the assumption that a, "product purchase" is equivalent to a "video view," because a user has to invest time into viewing a video on a particular topic. With this type of distance measurement, we can pull out the, "Principal Components" of the matrix via Principal Components Analysis and "Cluster" user behavior through an entire mathematical transform called, "K-Means Clustering," [7].

C. Performing Principal Component Analysis

Comparing every single person's click history to one another would be inordinately complex and difficult for a single analyst to understand - it would basically be like looking at an equation with hundreds or thousands of variables. To simplify things, we can break down those hundreds or thousands of variable equations into an equation of two or three variables. These two or three variables can then be viewed on a 2-dimensional or 3-dimensional graph, which is much easier to visualize. The process of reducing these N-variable equations down into a 2-variable equation is known as, "Principal Component Analysis," or PCA.

Put simply, PCA assigns, "weight," to all of the various variables in the matrix, ranking based upon, "lowest variance" in a particular direction, variance essentially meaning, "spread-outness." These sets of transforms are applied across all datapoint in a matrix, and the, "most important," meaning least spread out by, "direction" components are ordered and kept. In this situation, the ideal scenario is that the two or three, "components" that get, "kept" are much larger numbers and much more significant than all of the other components that were "normalized out." The derivation of what constitutes a single direction or component of a PCA is shown in Fig 9.

Once you are left with these two or three variables, you can then graph them out on a two-dimensional or three-di-

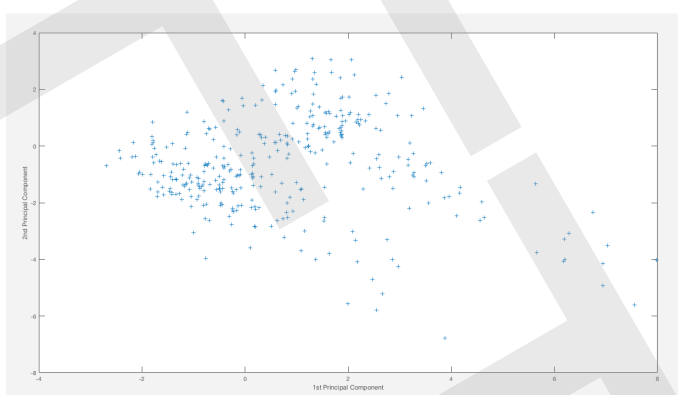

Fig 10. First and Second Principal Component Analysis of a selection of Session Viewer Data Matrix, graphed on X-Y space.

mensional space, which makes it much more readable by humans, as shown in Fig 10.

There is "hidden," information behind the main two dimensions represented in Fig 10., in the sense that, we transformed data from a highly complex space only used the first two principal components to create that graph in an X-Y plane. There are actually many, "components" of the above matrix calculated, which depending upon our original layout, may represent varying degrees of "importance," when mapping out and reducing data.

To graph out how important our first two selected principal components are related to every other component, we can graph all components out in a Pareto chart next to each other as shown in Fig 11. On the Y-axis we have, "percentage of explained variances" and on the X-axis we show the principal components. Ideally, the first two or three principal components would represent a maximum amount of variance

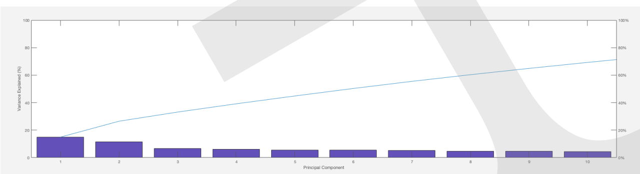

Fig 11. First and Second Principal Component Analysis of a selection of Session Viewer Data Matrix, graphed on X-Y space.

among all of the data we have mapped out. However, this is not always the case. The Pareto chart is a good place to examine the quality of a particular segmentation.

What we see in Fig 10., is that the first principal component accounted for about 18% and the second principal component accounted or about 15% of the overall variability of the original data set, showing which sessions were visited by which people. This means that the first two principal components account for 33% of the total variance, which might sound low, but since there is no actual reference point for what might be, "too low," for the purposes of improving performance as discussed in Section V., we withhold judgement and simply set it aside as a performance marker.

To help simplify what this means, we can consider the following: If all of the columns were of uniform height, there would be very little variability, and very little insight that reducing variables might give - the data would be, "spread out." In essence, the, "signal to noise" ratio is very low. However, if the sum of columns 1 and 2 were at close to 99%, then there is a risk of the model being too simplified - this would not have been likely to capture in nature and there may have been some problems on the input side.

There is a balance between making overly simplistic models of what's going on (or perhaps even introducing bias to artificially simplify a model, if we were to pick a different algorithm) in the world vs. having virtually no useful information to recommend a movement. Part of deciding what kind of Pareto chart works for a particular recommendation application may take time to, "harden," in terms of results further down the software pipeline. Indeed, even deciding whether PCA works at all may take time to understand whether the model is sufficient or may need more introduced bias to maximize variability, at least as a starting point. The important thing is to understand clearly that this is one point at which the performance of the algorithm can be monitored.

D. K-Means Clustering

Once the data is mapped out in 2 dimensions, the actual clustering process itself can be done mathematically in a fairly straightforward way. There are ways to measure the quality of the clustering, but that measurement can be done automatically. Naming the clusters and making sense of them may require a human touch.

A simplified way of explaining how clustering works, which is done by k-means clustering is that basically, the math runs through different numbers of cluster, and compares which number of clusters is the best by computing, "boundaries" based upon centers of each cluster, and whichever boundary works the best determines the proper number of clusters.

The boundaries are drawn and centers of each cluster are calculated by using the difference similarity metric discussed above in Section B and demonstrated again below in Fig 12.

$$\sum_{k=1}^{K}\sum_{i\in c_k}||x_i - m_k||^2 \qquad \text{Euclidean distance}$$

$$\min_{C,m_1\ldots m_k}\sum_{k}\sum_{C(i)=k}||x_i - m_k||^2$$

Fig 12. Iterative dimensionality computation of the distance between the centers for every single point, as calculated by Euclidian distance from the center. The enlarged problem is that clusters C are minimized for each K.

Fig 13. shows the results of a clustering computation for the above principal component analysis. The set of operations described above computers three clusters - green, red and blue. The performance of the different boundaries and cluster variations can be evaluated mathematically by measuring how many, "overlaps," resulting from each iteration, in what is called a silhouette chart. In the Fig 13, a variation using 4 clusters was not ideal because there were far more overlapping, improperly mapped points compared to 3 clusters, which had the minimum amount of overlap.

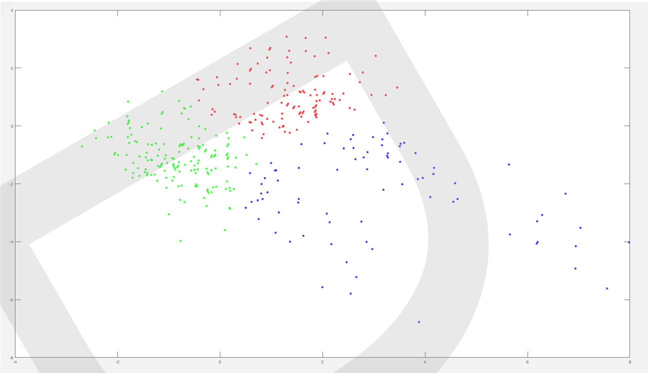

Fig 12. First and Second Principal Component Analysis of a selection of Session Viewer Data Matrix, graphed on X-Y space.

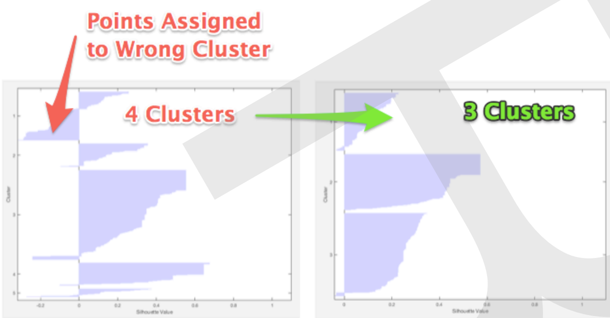

Fig 14. Silhouette charts can be used to optimize for the minimum amount of overlap between clusters and cluster boundaries.

E. Ascribing Meaning to the Clusters

After the clusters are mathematically determined, it is necessary to interpret the clusters and give some kind of insights to decision makers to help them take further action. This can be initiated through an inspection of the various elements in the map to first look at what types of users are present. Information about the users in terms of what sessions they were likely to view could be inspected prior to the clusters being given labels.

In the above example, we inspected user type based upon the content that they consumed. From here, we were able to generalize the above clusters and give it a human-readable, “tag” to help further understand the clusters.

The tags on the above chart are names only, and may also be tagged as any sort of name, much as what might be done during a market segmentation exercise. The user types shown may just as well have been given tags, “A,” “B,” “C,” to reduce bias. The tags we chose, shown in Fig 15., are based upon a cursory inspection of user profiles, and our judgement based upon the types of sessions viewed. Of course ascribing names to clusters in it of itself may seem against the purpose of Collaborative Filtering, which is to automatically tie together results, but as discussed earlier this paper, what we are looking for is a serendipitous output or recommendation to give to a marketer, who can then use subject matter expertise to act on the output.

Fig 15. Tags ascribed to each segment based upon a cursory inspection of user profiles, and our judgement based upon the types of sessions viewed.

What we actually do with these named clusters is discussed below in Section IV, Applying the Clustering Model and Testing Performance.

IV. Applying the Clustering Model and Testing Performance

A. The Control System Concept

"All models are wrong, but some are useful,” is a common aphorism in statistics. Ultimately, any model we build based upon video views for the purposes of generating superior online content is only as good as its ability to create more page views, more engagement, and ultimately more sales, per dollar marketing spend.

The original goal of this paper is to outline how algorithms can leverage video views to help decision makers to alleviate the problem of information overload. Taking a step back, let’s look at a general model which defines how we are solving this problem. The below graphic outlines various steps that we are taking, and how we can conceptualize the data being input, processed and output in a simplified, control-theory diagram as shown in Fig 16.

In Section III, "Performing the Cluster Analysis,” above we covered the left hand and upper part of this diagram, w, B, $\dot{q}$, $\int$ and q.

It’s important to consider at this point that there are several well known tools at a marketer’s or decision maker’s

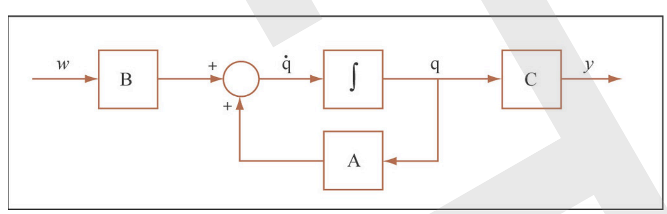

Fig 16. Control Theory Diagram representing our Recommender System and how it interacts with marketing actions.

TABLE IV
CONTROL THEORY DIAGRAM VARIABLES

Variable	Description
w	Raw training data, matrix of viewtimes per user
B	Weighted function, based upon time or human selection of video topic matter
q̇	Weighted input data
∫	Transfer function, clustering function
q	Weighted, clustered data, users and views tagged by cluster
A	Success function, showing which types of outputs created greater returns
C	further weighted, to make output data useful by marketing tools
y	finalized output

Table IV explains the variables of our Control Theory Diagram shown in Fig 16.

disposal which allows them to input a list of emails, and take an action (e.g., spend money) which gives them results in the form of clicks, page views, additions to shopping carts and ultimately purchases.

"A" may represent the input-output success model, and a further weighting factor which gets applied to q̇ to improve the overall model. A may represent weights and outputs from one or multiple tools which either currently exist or may evolve and emerge in the future, which take in emails as inputs, and create a marketing output. The following is a brief discussion on some examples of these marketing tools and how they may be used to create the, "A" feedback function above.

B. Facebook, Google Adwords, LinkedIn, and Other Platform Custom Audiences

Facebook was the first platform to allow the creation of custom audiences for the purposes of targeted ads, starting in 2013. Google, LinkedIn and all other major platforms have followed suit. These custom audiences can be built based upon inputs which an advertiser feeds into Facebook such as email, names, phone number or other information. Facebook will then build a, "lookalike audience," based upon a similarity scale which the advertiser selects. The idea of course is to create more effective content-audience matches and a higher return on advertising spend.

To create a "lookalike audience" an advertiser uploads their customer information as a "seed audience" which the platform such as Facebook compares to its entirety of user profiles to find the commonalities. The platform then generates a, "lookalike audience," abstraction, which the advertiser can then opt to send paid advertising to. The objective of creating a lookalike audience as opposed to an audience created based upon manual parameters is that the automated method should hypothetically be better at finding highly-qualified customers based upon user data and behavior who previously would have been difficult to identify and reach.

Fig. 17 shows some of the input parameters asked by Facebook on its, "Custom Audience" interface, which includes email. The input rows uploaded are known as a, "seed audience," while the output is a, "lookalike audience." How Facebook and other social media platforms create lookalike audiences is opaque, but based upon existing studies, the audiences are thought to be built primarily through clustering and similarity metrics much like those demonstrated in this paper, however on a much larger scale.

	A	B	C	D	E	F	G	H	I	J
1	email	email	email	phone	phone	phone	madid	fn	ln	zip

Fig 17. Facebook custom audience input parameters.

The input or seed audience used to create lookalike audiences requires a minimum quantity of inputs. At the time of writing, Facebook requires a minimum of 100 inputs. Other platforms may have different requirements.

C. Email Marketing

Email marketing for some industries is alive and well today and can often have a very high return on investment if done well. Email marketing can function much like social media feeds, albeit direct to a messaging inbox rather than a general feed. However, there are barriers to email marketing such as an ever-increasing chance of being captured and identified as spam by automated filters. The only real way to successfully reach a customer's inbox is to provide delightful, engaging, informative and useful content to that user.

What constitutes delightful, engaging, informative and useful varies from email recipient to email recipient of course. Email readers are consistently trying to achieve differing goals over time and so being able to break up email readers into smaller groups and feed more effective content based upon common goals through experimentation is a way to, "do the best for one's customers," and keep emails useful. Our clustering model discussed above, combined with a memetic model which ranks topics per cluster, as shown in Fig 18., can be used to suggest new email campaigns on a periodic basis in order to keep content fresh.

We can build a, "memetic ranking" model among clusters of data to show which titles of various videos found the

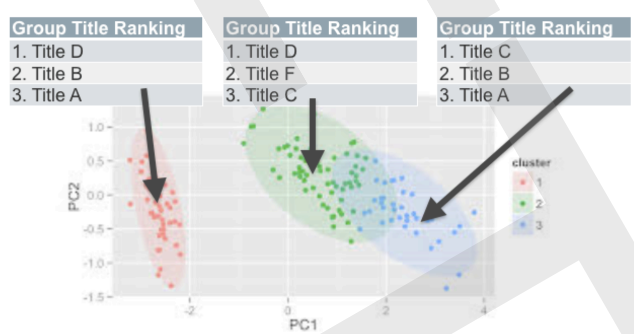

Fig 18. Example graph group word or title ranking per cluster, with clusters built on top of an already created PCA.

most popularity within a particular cluster. This type of title ranking can help supply a serendipitous answer to the question of, "What to write," or, "what to produce," in terms of predicting what topics may confer the greatest amount of open rates or click rates. The function which dictates the below graph could simply be topics ranked by popularity for a given cluster.

D. Blog Posts, Further Topic Selection & SEO Analysis

Search Engine Optimization (SEO) is the, "long, slow," version of pulling in marketing leads and building sales based upon topics and blogs written over time. SEO, much like email marketing is driven based upon topic interest over time. The same types of memetic models could be formed as with Email marketing to suggest what topic areas to, "enter into," based upon memetic ranking. Because writing blog posts and in particular quality blog posts can be expensive, being able to pick the right topics is critical to build an appropriate SEO foundation. This same logic applies to the creation of video and other high forms of content.

V. Adapting the Clustering Model

A. Testing Performance Against Flat Audience

Of course, building mathematical models for their own sake is merely a hobby. The usefulness of a model can only be taken as far as the difference in results it may produce from a baseline. The performance of a predictive recommender system may be evaluated by running parallel tests, and setting clear performative metrics to evaluate the parallel tests against one another. It is never good enough to put a model in place and to say, "trust me, we are better off than we would have been." There needs to be a stronger demonstrable comparison to point toward this conclusion. Considering the control systems model which we proposed above:

The function, "A" is a set of model ranking systems which help determine what the action should be on the next loop. The line of inputs, "w" through outputs "y" are data points that any data processing department in an organization working with video views and website clicks may already have on hand. However the box, "∫" may represent a set of either clustered or non-clustered functions to attempt to augment the performance of the overall system from "w" to "y." The "A" function could be built to filter out the results of both clustered and non-clustered models being used within "∫1" and "∫2" with different functions being built and run in parallel to each other, one with clustering and one without.

Should the results of ∫1 provide a higher ranking based upon necessary parameters than ∫2 in terms of evaluating performance of the "y" given "w," then A may weight ∫1 over ∫2 in the next iteration, or vice-versa. If it turns out that ∫2 significantly underperforms ∫1 over time, then ∫1 would come to dominate the overall control system model. An abstract explanation of how the, "A" evaluation function may be built is shown in Fig 19. Ultimately, if the clustering technique does not work for a particular application, given the above evaluation weighting function, its usage would reduce down to 0.

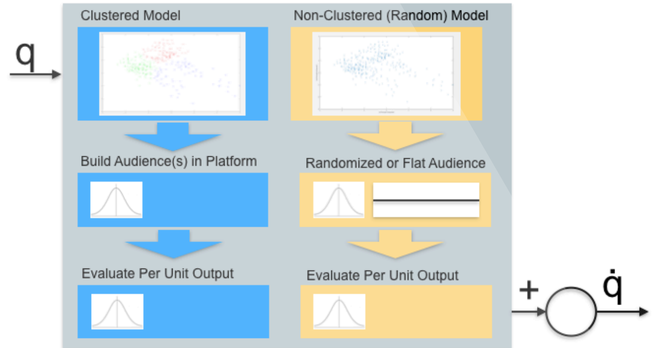

Fig 19. Abstract image explanation of evaluation test comparing clustered models to non-clustered (random) models and adapting Weighted Input Data $\dot{q}$ to improve overall output.

B. Weaknesses of Clustering Approach

The advantage of Collaborative Filtering and using Principal Components Analysis to help, "map things out," is that it makes complex things easier to understand and can help with faster decision making. The disadvantage is, you may end up with a model which unfortunately has very little in the way of recommending anything, "useful," or, "meaty" in that it points toward a strong signal in the marketplace, or able to create a return on investment. This disadvantage is also termed as, "overspecialization," or "data sparsity," with more disadvantages discussed in table V [6].

C. Weaknesses of Segmentation Recommendation

Assuming that our clustering model creates, "useful," output in terms of data being sufficiently dense as measured by Pareto charts on the principal components and having well-defined segments as measured by silhouette diagrams, there is a disadvantage in applying labels to segments and linking segments to topic matter, which is that follow-on activities may not be effective in it of themselves. Ultimately,

TABLE V
WEAKNESSES IN CLUSTER MODELING

Weakness	Description
Data Sparcity	Insufficient or flat data per parameter. If a marketer wanted to zero in on a particular topic, there may not be enough data on that one topic to generate rankings related to that topic.
Cold Start	Insufficient data at start of usage of new program.
Scalability	Typically a problem for creating many recommendations on an individual basis, computing resources are less of an issue if recommendations are being presented to a single team as a complete computation can be performed, "in the background," rather than instantly on a per user basis.
Synoniminity	This problem exists more for clustering similar terms which might be synonymous. This is not an issue for serendipitous market segmentation.

Table V explains various weaknesses in the usage of Cluster Modeling for Recommendations.

subject matter expertise in a particular market segment is always needed to create effective advertising overall.

VI. Conclusion

Marketers, data scientists and generally those tasked with characterizing entire industries and sectors have an increasingly complex task of understanding an increasingly complex and dynamic, changing world, to make better decisions which drive profitable results. Our above suggested clustering methodology is one specific approach that represents an overall generalized way to filter, prioritize and efficiently deliver relevant information to different audiences en masse, while keeping track of the return on investment of said clustering methodology. Usage of clustering has the potential to improve outcomes, but must keep in mind the parameters used within the clustering method, and ideally the performance of the clustering model must be tested against a non-clustered approach to ensure that clustering in it of itself drives results.

Acknowledgment

While this paper was not written as a part of an academic institution, Mr. Delaney would like to acknowledge the influence of his former professor, Dr. Vladimir Cherkassky for encouraging a philosophical approach to data science as well as Mr. Steve Mase for encouraging a business minded approach to data science and his wife, Andrea Salazar, for supporting and encouraging his work through the years.

References

1. Greg Linden, Brent Smith, (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Computer Society IEEEE Internet Computing 1089-7801/03
2. Sara Dolnicar, (2002). A Review of Unquestioned Standards in Using Cluster Analysis for DataDriven Market Segmentation. Wollongong, Australia: Faculty of Commerce-Papers, Page 273 University of Wollongong Research Online
3. Ghulam Shabbir, Khan Niazi, Javaria Siddiqui, Burhan Ali Shah, Ahmed Imran Hunjra (2012). Effective Advertising and its Influence on Consumer Buying Behavior. European Journal of Business and Management www.iiste.org ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online) Vol 3, No.3.
4. Lawrence Angrave, Zhilin Zhang, (2020). Who Benefits?: Positive Learner Outcomes from Behavioral Analytics of Online Lecture Video Viewing Using ClassTranscribe. Proceedings of the 51st ACM Technical Symposium on Computer Science Education Pages 1193–1199
5. Sheena Leek, George Christodoulides (2012). A framework of Brand Value in B2B markets: The Contributing Role of Functional and Emotional Components. Industrial Marketing Management Volume 41, Issue 1, Pages 106-114.
6. F.O. Isinkayea, Y.O. Folajimib, B.A. Ojokohc, (2015). Recommendation Systems: Principles, Methods and Evaluation. Cairo, Egypt: Pages 261-273 Egyptian Informatics Journal.
7. Chris Ding, Xiaofeng He (2004). K-means Clustering via Principal Component Analysis. Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004.

Patrick W. Delaney (B.S.E.E. 2008 UMN) is the founder of a nationally recognized Internet of Things Conference, IoTFuse, as well as a premiere Medical and Healthcare conference, MedFuse and other conferences. Mr. Delaney's career has spanned from electronics design to software development, to semiconductor sales for a Global Fortune 20 (Samsung) and marketing at another Global Fortune 20 (United Health Group).