Patterns of Philanthropy: Using Pattern Mining for Predictive Analysis in Advancement and Fund Raising

Klaus Mueller

Akai Kaeru LLC, Stony Brook, NY USA mueller@akaikaeru.com
Computer Science, Stony Brook University, NY, USA

Eric Papenhausen

Akai Kaeru LLC, Stony Brook, NY USA epapenha@akaikaeru.com

ABSTRACT

To support their academic mission universities and colleges have become increasingly dependent on raising capital via advancement channels. As these institutions compete for the attention and the funds of donors advanced data analysis is a key to success. We present a method and system that can identify specific groups of donors which share sets of demographic, academic, and other features. This information can then be used to shape specific fund raising efforts and evaluate the expected profitability of these. We demonstrate our system using an advancement dataset we obtained via collaboration with a public university.

1 Introduction

Advancement offices of universities have access to a wealth of data, characterizing alumni and previous donors by many features, or *attributes*. Identifying groups of individuals in these high-dimensional attribute spaces can reveal valuable insights on the potential of other, similar individuals to also make donations, predict the size of these donations, and help shape the message used to gain their attention. In this paper we report on a study we conducted using a state-of-the-art machine learning approach and system we developed recently. Our method and system can identify the specific characteristics of donors in a large database and predict their prospects of future donations along several instruments.

2 Methods

Our study uses a dataset we obtained via collaboration with a public university with over 50,000 undergraduate and graduate students and over 24,000 faculty and staff. The dataset has 168 attributes covering demographic and academic information as well as donations for 2,054 donors (1,093 managed and 961 unmanaged). Our pattern mining engine looks for regions in this feature space that are occupied with similar donors that all respond in a similar way to a given target variable of interest, here the type of donation.

Each pattern of similar donors forms what is called a *subspace* [1]. It is a subpopulation of donors that fit inside a low-dimensional hypercube with well-defined value ranges of the features that describe the subspace. This property, and the fact that these subpopulations are typically rather low-dimensional, even when the overall feature space is not, makes them easy to understand and explain [2]. We exploit this property for the study presented here, and note that while deep neural networks, random forests, etc. also learn low-dimensional representations, these are not easily described in terms of their native attributes.

Concretely, given a dataset with attributes $\{A_1 A_2 \ldots . A_m\ P\}$ with P being an attribute of interest, such as the amount or frequency of a donation, the goal of pattern mining is to find a hypercube (or pattern) consisting of constraints of the form $A_i \in [v_l, v_r]$ for $i \in [1 \ldots m]$ (for example, *age* > 45, *degree* = PhD), where the points within the pattern are "interesting". A pattern of donors will be considered interesting if the probability of a specific type of donation (e.g. lifetime endowment, lifetime planned gift, etc.) is significantly higher than the overall probability. The definition of what constitutes a consistently interesting pattern is based primarily on statistical hypothesis testing. For numerical attributes we use the Mann-Whitney test [3] to account for the often non-parametric nature of the data while for a binary target attribute we use the $\chi 2$ test for independence.

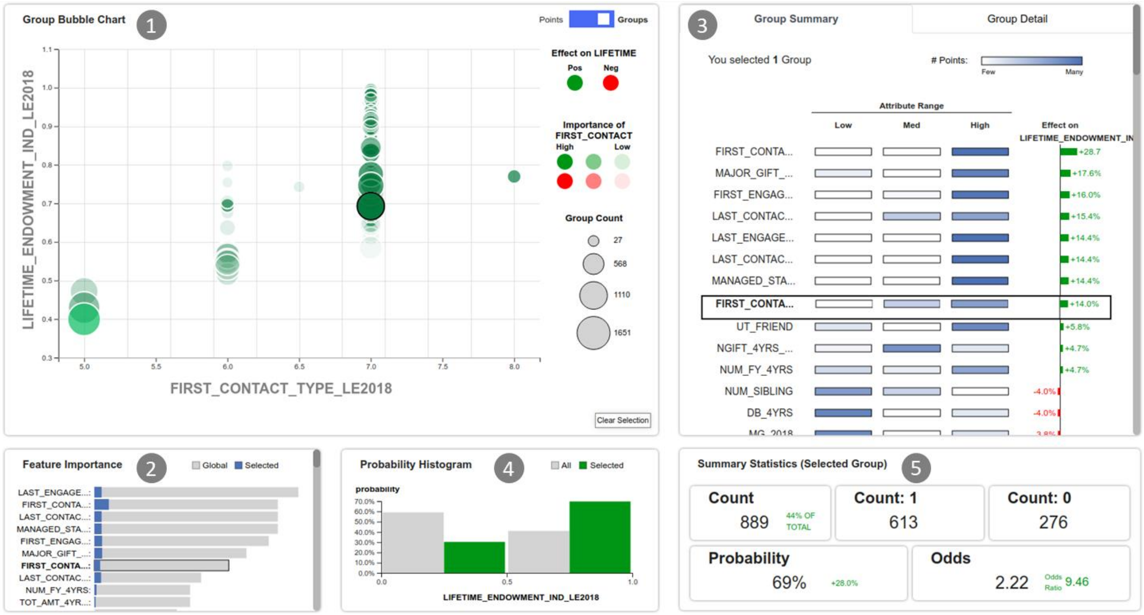

Figure 1: Visual interface showing the impact of FIRST_CONTACT_TYPE on LIFE_TIME_ENDOWMENT_IND. The values on the x-axis are Email (5.0), Other (6.0), Scheduled Visit (7.0), and Phone/Video Call (8.0), See the main text starting Section 3 for a description of the interface components labeled 1-5.

Extracting the patterns requires extensive search. Our pattern mining algorithm is based on the FP-growth algorithm [4]. One of its main advantages is that it only requires scanning the full dataset twice throughout the mining process. This makes it more scalable in contrast to many other pattern mining algorithms in which the dataset must be repeatedly scanned. Given many additional optimizations, our pattern mining algorithm can analyze even large datasets within a couple of minutes.

3 Visual Interface: Introduction by Example

Fig. 1 shows the interactive visual dashboard interface by which analysts can access the results produced by our pattern mining engine. All data refer to the year 2018 and earlier. For this first demonstration let's follow Bob, an advancement analyst at a large state university. His goal is to explore the attractive fund raising instrument of lifetime endowment. This is captured in the variable LIFETIME_ENDOWMENT_IND which is an indicator set to 1 if a donor committed to a lifetime endowment in 2018 or earlier; else it is set to 0. After selecting this variable the system performs the pattern mining and populates the various charts of the dashboard. Bob's first focus is the Feature Importance chart (panel 2); it shows a bar chart of features where the length of a bar indicates the predictive strength of the associated variable for the given target variable, LIFETIME_ENDOWMENT_IND (our system obtains these values via Shapley analysis [5]).

In the Feature Importance chart Bob observes that the variable FIRST_CONTACT_TYPE is one of the most predictive variables (it is ranked #7 and has a fairly long grey bar). This variable piques Bob's interest since he would like to plan the best way in which to contact donors inclined to step up to a lifetime endowment. He clicks on the FIRST_CONTACT_TYPE bar after which Panel 1, the Group Bubble Chart, is populated. The Group Bubble Chart shows the relation of the mined patterns with respect to the target variable LIFETIME_ENDOWMENT_IND and the predictor variable FIRST_CONTACT_TYPE. Each bubble is a pattern – a group of past donors – where the size of a bubble indicates the size of the group's subpopulation. Bob pays particular attention to the bubbles colored in a saturated green as these are groups of donors where FIRST_CONTACT_TYPE bears a significant predictive value.

The values on the x-axis are Email (5.0), Other (6.0), Scheduled Visit (7.0), and Phone/Video Call (8.0). Bob notices that Email (5.0) seems rather ineffective and that while Phone/Video Call (8.0) looks better there is only one rather small group that has apparently responded to it. On the other hand, Scheduled Visit

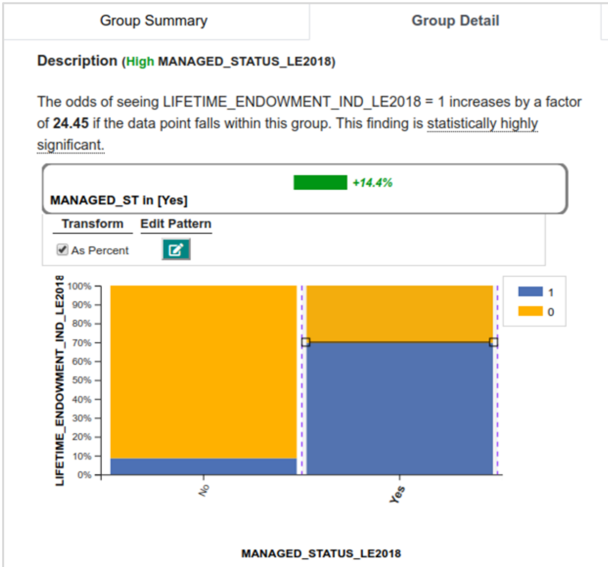

Figure 2: Group Detail chart for the MANAGED_STATUS attribute in Fig. 1 (obtained by selecting the tab in panel 3).

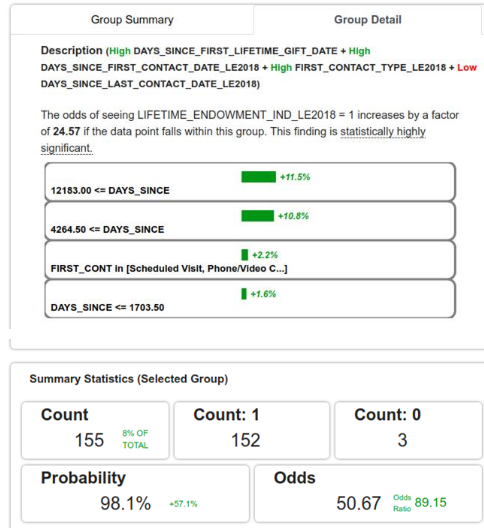

Figure 3: Group Detail and Summary Statistics panels after clicking the top most bubble in the Scheduled Visit (7.0) column of Fig. 1.

(7.0) seems to have more promise. There are multiple groups placed at different positions along the y-axis. The higher up the bubble is placed the higher the proportion of donors where LIFETIME_ENDOWMENT_IND is set to 1.

Bob selects the large solid green bubble at y ≈ 0.7. There are other bubbles with a higher predictive probability but this one is the largest (i.e. has the most donors) and so Bob decides to explore it further. This renders the dashboard as shown in Fig. 1 (the blue filled regions in the Feature Importance bars (panel 2) now show the respective feature importance of the group represented by the selected bubble). The Summary Statistics (panel 5) shows that the group is made up of 889 donors; of these, 613 donors have LIFETIME_ENDOWMENT_IND set to 1 and 276 donors have it set to 0. The predictive quality of FIRST_CONTACT_TYPE = Scheduled Visit to lead to a LIFETIME_ENDOWMENT is therefore 613/889 = 0.69 (which corresponds to the placement of the bubble on the chart's y-axis).

The Probability Histogram (panel 4) confirms that the pattern's population (green bars) has a higher probability of LIFETIME_ENDOWMENT_IND =1 than the overall population (grey bars). Next, Bob focuses on the Group Summary Chart (panel 3). In the feature list he sees many significant attributes (indicated by the longer green Shapley value bars on the right) and he also notices that their values all fall into the high attribute range. The prevalence of the dark blue boxes tells Bob that these attributes are all highly correlated. These are attributes like FIRST_CONTACT_DATE, LAST_ENGAGEMENT_DATE and so on; Bob knows that high values there indicate that these donors fall into the general category of 'managed donors' – donors who are on the advancement office's special focus list.

Bob confirms his initial assessment when he notices that the variable MANAGED_STATUS is also high on the pattern's feature list and that its values predominantly fall into the high range level; it means that they have been set to 1. He clicks on the Group Detail tab in panel 3 to learn more about the participation of this donor type in this donor group. This brings up the information shown in Fig. 2. Bob confirms that indeed about 70% of the managed donors are lifetime endowment donors (blue region, right bar), while only about 8% of the unmanaged donors are lifetime endowment donors (blue region, left bar). The narrative in Fig. 2 states this in natural language: if a donor is managed then the odds of the donor making a lifetime endowment is increased by a factor of over 24 (and this finding is statistically highly significant). Given these persuasive findings, Bob plans to focus his further analysis on the managed donors.

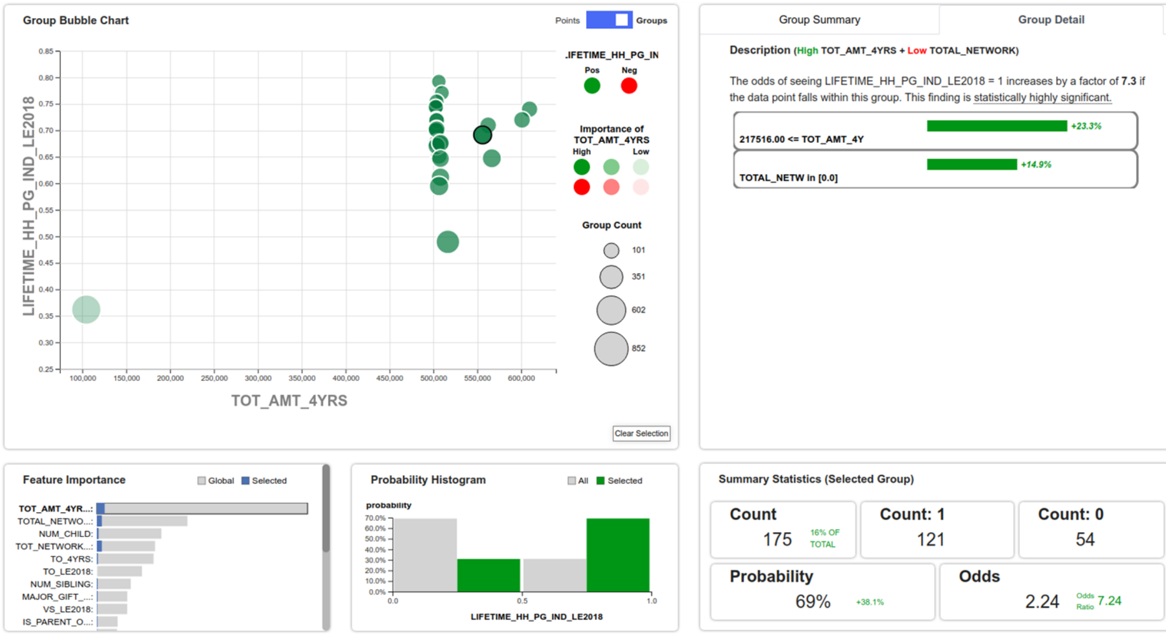

Figure 4: Dashboard visualizing the identified subpopulations of managed donors with respect to their propensity of making a planned gift. The current view arranges the groups in terms of the total amount of gifts these households have made over the past 4 years (TOT_AMT_YRS). The analyst has selected the group bubble marked by a thick dark boundary and the associated information is shown in the various charts.

Zoe is a colleague of Bob and has been participating in the analysis. She is wondering about the other bubbles stacked above the bubble Bob just explored as they give rise to even higher probabilities of LIFETIME_ENDOWMENT_IND. Adventurous, Zoe clicks on the top most bubble. She glances at the Summary Statistics (see Fig. 3, bottom) and notices that the group is much smaller, about 20% of the group the team explored before. But the probability of making a lifetime endowment of this small group is close to 100%. She clicks on the Group Detail tab (Fig. 3 top) and learns that this donor group seems to have enjoyed high attention for a long time; the time since the first contact and the time of the first gift are both high, while the time since the last contact is small. Apparently managing the donors really pays off. But nevertheless, the group is rather small. Further analysis reveals that they are a select group of truly wealthy donors who have been in the system for a long time.

4 Selected Case Studies

In this section we demonstrate our system via a set of further analysis tasks. We continue to follow Bob and Zoe in their quest to learn more about the particularities of their university's existing fund raising efforts with the goal of identifying promising (and not so promising) strategies to increase the yield moving forward. Per the prior findings they focus on managed donors for the time being.

4.1 Identify Managed Donors Who Might Make a Planned Gift

Bob knows that goal #1 in effective fundraising is to know your most promising targets. Continuing with the general population of managed donors, Bob decides to hone into the indicator variable LIFETIME_HH_PG_IND which is set to 1 when a household has made (or is anticipated to make) a planned gift. Planned gifts are typically difficult to predict as they often occur in a will, after the donor has passed and without prior announcement [6]. Predictive analysis based on historical data can be of immense value in assessing the propensity of a certain type of donor to make a planned gift.

Specifying LIFETIME_HH_PG_IND at the target variable gives rise to the dashboard shown in Fig, 4. Bob observes quickly, in the Feature Importance chart, that the most predictive indicator for this target is the

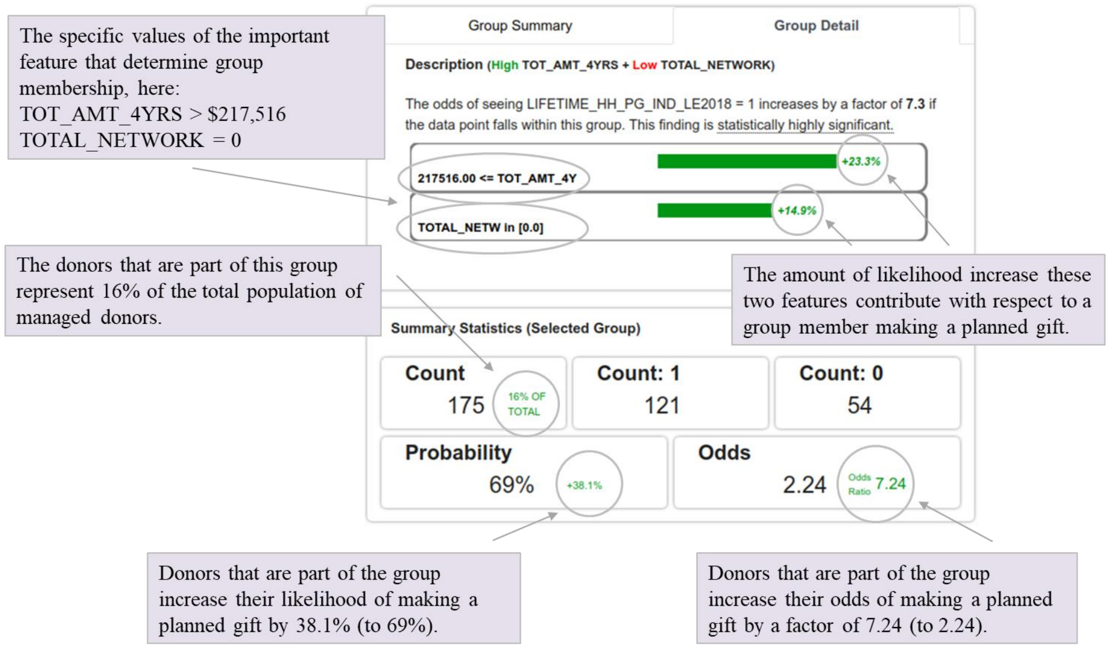

Figure 5: Annotated Summary Statistics and Group Detail panel.

total amount of gifts these households have made over the past 4 years, as captured in the variable TOT_AMT_4YRS. The next most important indicator is TOTAL_NETWORK – the size of the donor's family, summing the variables NUM_CHILD, NUM_SIBLING, OTHER_CHILD. The variables following on the list also all have something to do with family network and prior gifts. Bob has a feeling that he is on the right track.

He clicks on the top-most variable TOT_AMT_4YRS and the various charts in the dashboard are populated (in fact the system automatically populates the charts by the top most variable when a new target is selected). All identified groups where TOT_AMT_4YRS is relevant (colored in a saturated green) are in excess of the $500k mark; this value indicates the median $ amount of the donations made by the donors in a group. Bob picks a group that is on the high end.

The Summary Statistics and the Group Detail panel offer specific numerical detail on this group, which represents about 16% of the overall population of managed donors. Fig. 5 annotates the various quantities revealed in these panels. While the probability of the overall population of managed donors to make a planned gift is just around 30.9%, being a member of this group adds 39.1% to this, more than doubling it to 69%. Likewise, their odds of making a planned gift is 121/54 = 2.24; this is a statistically highly significant increase by a factor of 7.24 over the general population of managed donors.

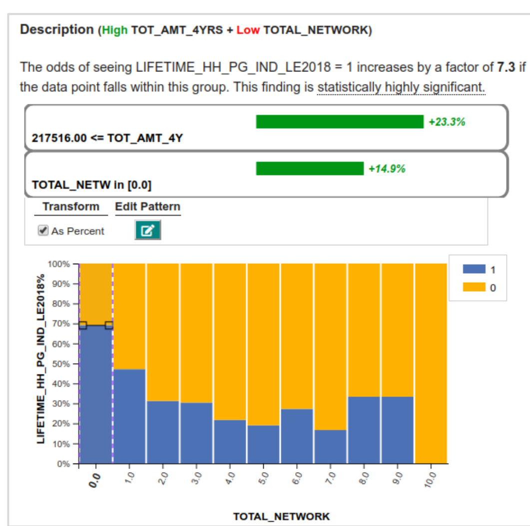

Figure 6: The detail bar chart of the TOTAL_NETWORK variable for the group selected in Fig. 4

Having uncovered this important subpopulation of donors Bob turns now to identify the key characteristics if this group. He examines the Group Detail panel and sees two features tabulated there: TOT_AMT_4YRS and TOTAL_NETWORK. They contribute 22.3% and 14.9%, respectively, to the overall increase in probability (38.1%) that a group member makes a planned gift in the

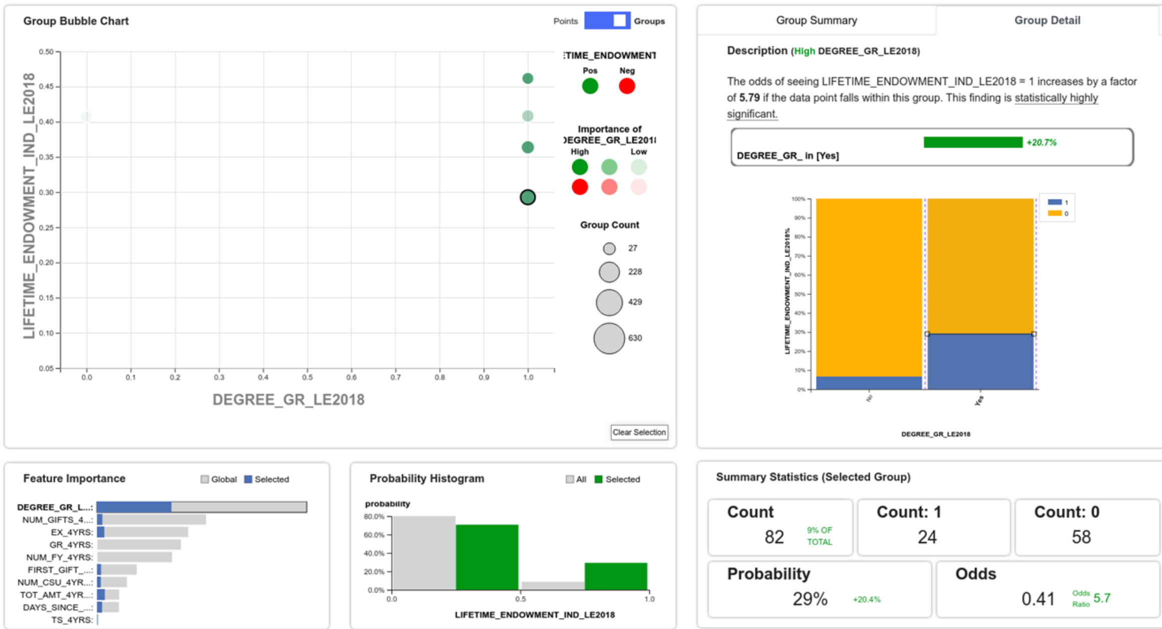

Figure 7: Dashboard visualizing the identified subpopulations of unmanaged donors with respect to their likelihood of making a lifelong endowment. The top feature is DEGREE_GR which are grads of the Business School; other schools do not appear. The analyst has selected the group bubble marked by a thick dark boundary and the associated information is shown in the various charts. It appears that this subpopulation increases the likelihood from 8.6% to 29%.

future. But most importantly, the panel also reveals the specific values of the important features that determine group membership, in this case. TOT_AMT_4YRS > $217,516 and TOTAL_NETWORK = 0. Here, the minimum 4-year total donation a donor in this group has made is $217,516 and the donor typically has no children, siblings, etc. Clicking on TOTAL_NETWORK brings up a bar chart that gives more information about the probability of each network value (see Fig. 6). Bob learns that while TOTAL_NETWORK = 0 has the highest probability, TOTAL_NETWORK = 1 is also somewhat useful. He would need to explore other variables in the Group Summary feature list to find out the specific type of network member associated with this, such as child, sibling, etc. These are the specific characteristics our pattern mining algorithm is able to expose.

Examining other donor groups in the Bubble Chart by simple mouse clicks reveals that other groups have similar but not identical characteristics. While the main themes follow the overall Feature Importance chart for the population of managed donors, drilling into individual subgroups allows Bob to provide very specific and nuanced recommendations to address each group of donors. He passes these recommendations on to the advancement office's marketing team.

4.2 Identify the Most Charitable Unmanaged Donors

While Bob's focus has been mostly on the managed donors, Zoe has become interested in the unmanaged ones. She switches the focus to this donor population and generates the dashboard shown in Fig. 6. It returns back to the LIFETIME_ENDOWMENT_IND variable as a target. Zoe immediately notices that DEGREE_GR tops the Feature Importance list by a wide margin, followed by more general attributes and another GR-type feature. GR stands for Business School and DEGREE_GR are donors who received an undergraduate or graduate degree from the Business School.

It appears that graduates from the business school are the most valuable prospects for advancement efforts. While the database also contains fields for the College of Fine Arts, the School of Engineering, the School of Social Work, and many others, only the Business School degree holders are involved in forming patterns associated with a promise of future lifetime endowments. However, even though the probability is not overly high, as is revealed when clicking on the most-solid pattern, it is still much higher (29%) than that of the

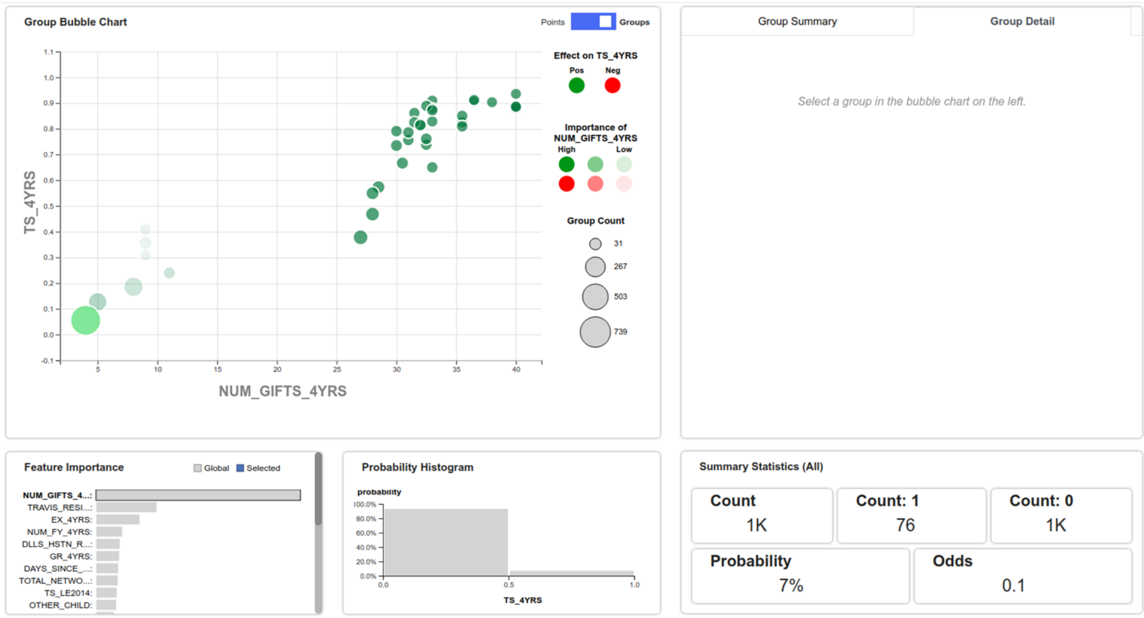

Fig 8: Dashboard visualizing the relation of radio station donations (TS_4YRS) and the number of university gifts.

overall unmanaged population (8.6%) as is shown in the Probability box of the Summary Statistics panel and is also visualized in the bar chart in the Group Summary panel.

4.3 Exploring the Donations to the Campus Radio Station

Bob is a big fan of the campus radio station. He actually DJ'd there when he was a student at the university. He knows that donations are the prime source of funding of the station, similar to most non-commercial radio stations. He fires up the pattern mining program (for the managed donors) and sets the target variable to TS_4YRS which gauges the amount of donations to the campus radio station over the last 4 years. Fig. 8 shows the dashboard that is generated using the top ranked feature in the Feature Importance list, NUM_GIFTS_4YRS, which is the number of gifts a donor has made over the past 4 years. We clearly observe a linear relationship – donor groups who are more generous to the university also give more to the campus radio station.

While this is interesting, it is too general of a finding. In the next sections we will look over Bob's shoulder as he explores some of the individual groups. Bob begins this journey by clicking on the TRAVIS_RESIDENT feature which changes the dashboard to the one shown in Fig. 9. He observes that Travis residents (TRAVIS_RESIDENT = 1) give far more to the campus radio station than non-residents (TRAVIS_RESIDENT = 0). The bubbles for the latter are also very faint which means that for these groups county residency is not a relevant feature. Bob of course knows that Travis County is the location of the university and so the significance of residency in Travis County makes perfect sense.

4.3.1 Exploring the largest donor group with lower median donations to the radio station

Bob starts the group-wise exploration with the largest bubble, placed lowest in the stack where TRAVIS_RESIDENT =1 (see Fig. 9). This group has 278 members and is characterized by just a single property, i.e. being a Travis County resident; it is the only bar shown in the Group Detail panel. It means that any managed donor who is a Travis County resident will give to the campus radio station at a likelihood of 24%. This is 17% greater than the likelihood of non-residents to contribute, as can be gleaned from the bar chart and also the Probability box in the Summary Statistics panel.

4.3.2 Exploring a mid-sized donor group with higher median donations to the radio station

Bob is now curious about the other bubbles. They are smaller and so they will correspond to smaller groups. He clicks on the bubble about halfway up the stack. The (partial) dashboard and the revealed statis-

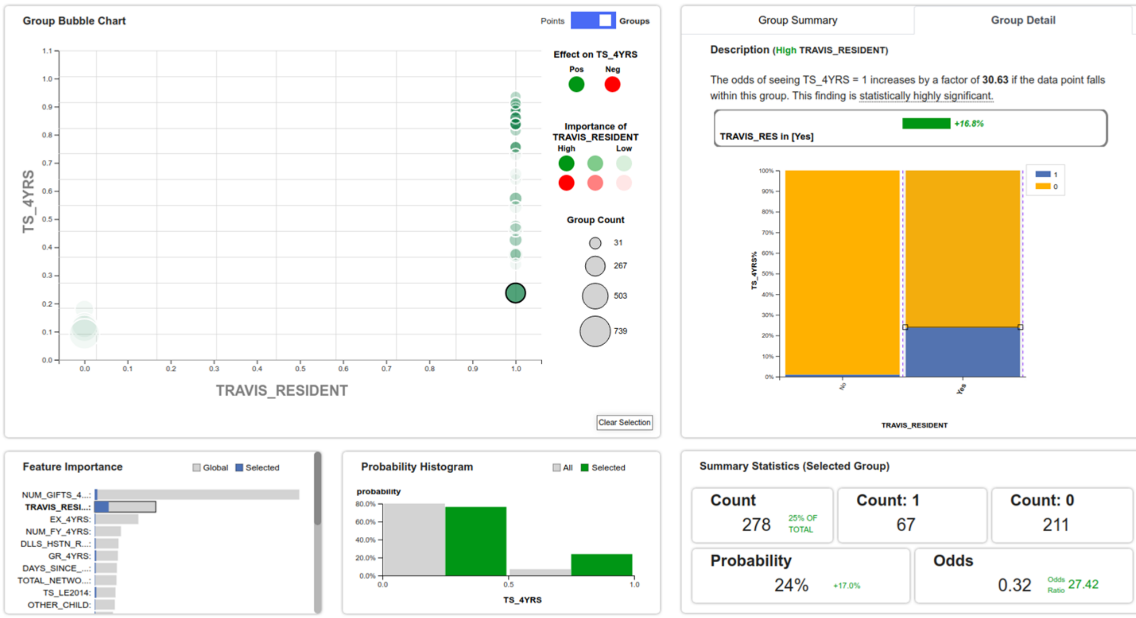

Fig 9: Dashboard visualizing the relation of radio station donations (TS_4YRS) and Travis County residence. The charts show the statistics of the bottom most bubble where TRAVIS_RESIDENT =1.

tics are shown in Fig. 10a. Bob observes that the group indeed has fewer members than the first group, 68, but that the likelihood of giving has risen to 57.4%, more than twice than the initial group. This group seems to be a lot more generous! Bob also notices that this group is more tightly defined. There are now two features represented by bars: NUM_GIFTS_4YRS and TRAVIS_RESIDENT.

Upon closer inspection, Bob learns that the most important feature for this group is NUM_GIFTS_4YRS; the number of gifts the donor has awarded the university over the past 4 years. It contributes 32% of the overall group likelihood. The number next to the bar indicates a minimum of 19 gifts. To gain more insight on the actual distribution he clicks on the bar which brings up the scatterplot shown in Fig. 10b (it would show below the bar but we have displaced it for presentation purposes). Each point in the scatterplot is a data point, a member of the group. The TS_4YRS attribute is a binary variable, set to 1 if a person gave to the radio station in the 4 years or 0 if not. To better show their distribution we have varied the y-coordinate of each point slightly, using a randomized process called jittering. The grey box contains the points for

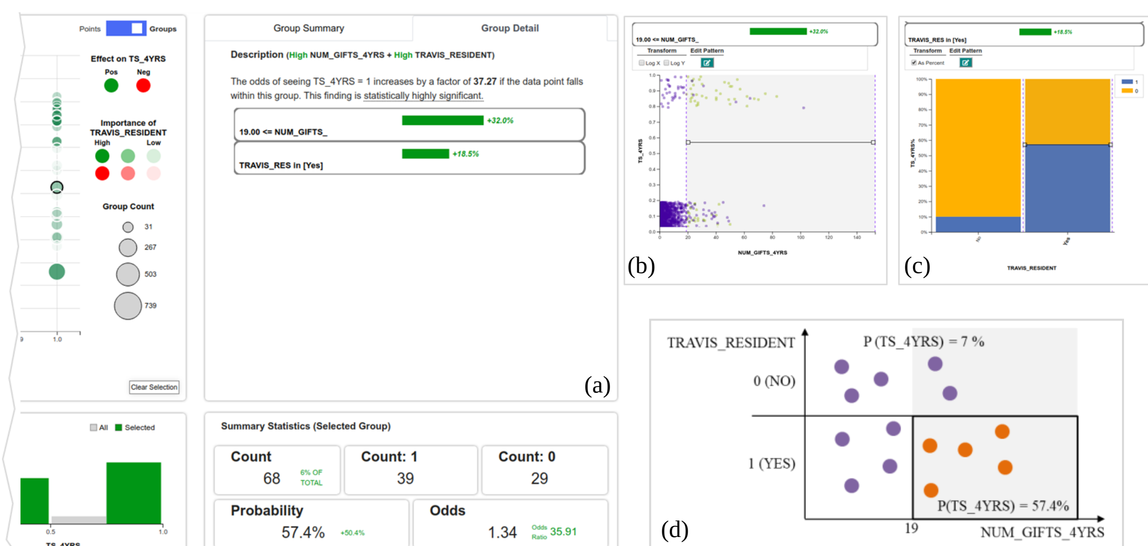

Fig 10: Dashboard visualizing the relation of radio station donations (TS_4YRS) and the number of university gifts. (a) Clipped main dashboard, (b, c) plots for the two features describing the pattern, (d) hyperbox illustration.

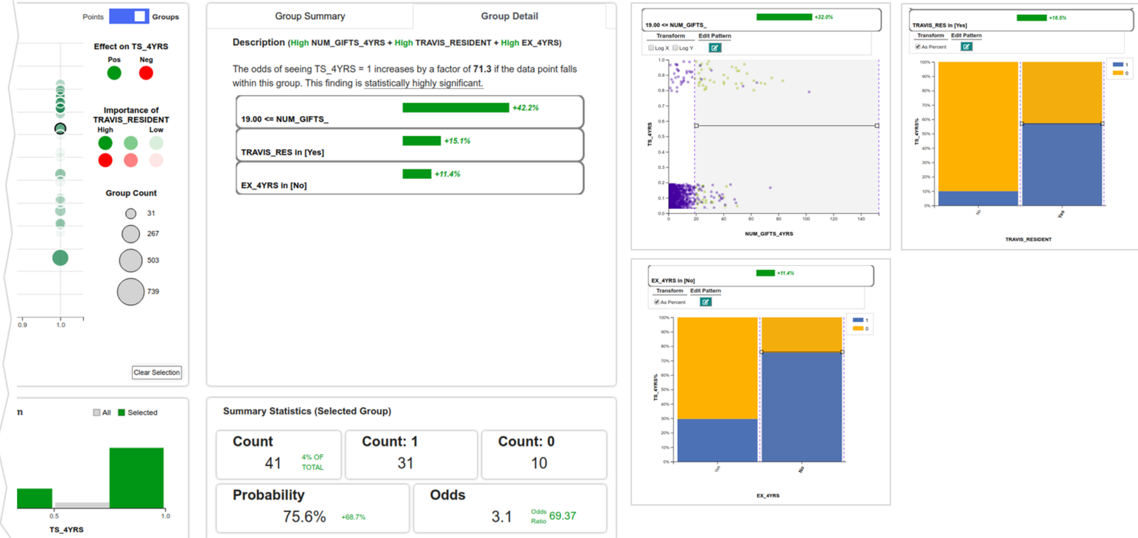

Fig 11: Dashboard visualizing the relation of radio station donations (TS_4YRS) and the number of university gifts. A group further up in the stack of bubbles has been selected. This group is defined by three features: NUM_GIFTS_4YRS, TRAVIS_RESIDENT, and EX_4YRS. Plots for each are shown on the right, ordered left to right, top to bottom.

which the average of TS_4YRS is significantly higher than the overall population average (calculated over all managed donors and tested via the $\chi 2$ test for independence, see Section 2). The yellow points are donors inside the group while the purple points are donors outside the group. It can be clearly seen that there is a clear majority of yellow points on the top of the plot where TS_4YRS = 1. Overall, Bob appreciates this plot as it gives him more insight on the donor profile than just a single number. For example, he sees that the more generous donors, those far to the right, also give to the radio station. Further, the plot also clarifies that donors who make no contributions to the radio station are few and overall on the less generous side, as seen by the rather few yellow points in the lower left corner of the grey box.

Bob notices that there are still some purple points in the scatterplot. These points are actually outside the pattern's hyperbox. This is illustrated in Fig. 10d where the grey box contains all points for which NUM_GIFTS_4YRS $\geq$ 19 (for ease of drawing the illustration neglects the third axis TS_4YRS which would go out of the plane). The orange points are points inside the pattern while the purple points are outside of it. Recall that this is a pattern defined by two attributes. The second attribute is TRAVIS_RESIDENT which is shown here as the y-axis. Bracketing the pattern by the line where TRAVIS_RESIDENT = 1 isolates the orange from the purple points and defines the pattern completely (captured in the solid box). It is worth noting that while this is a simple 2D box case, there is no limit on the number of dimensions that might be needed to establish statistical significance of a pattern, resulting in a N-D hyperbox. Our pattern mining algorithm finds these hyperboxes automatically. Fortunately, in practice the hyperboxes tend to be rather low-dimensional, rarely more than 4D, and so are easy to explain.

Adding the condition TRAVIS_RESIDENT = 1 to the pattern description adds another 18.5% of probability in favor of a donation to the radio station. The bar chart for the resulting distributions is shown in Fig. 10c. Note that these distributions assume that the first condition NUM_GIFTS_4YRS $\geq$ 19 has been met.

4.3.3 Exploring a smaller donor group with even higher median donations to the radio station

Bob moves up the stack of bubbles and clicks on the group indicated in Fig. 11. This is a group with three conditions, a 3D hyperbox. As indicated by the three bars in the Group Summary panel, the three conditions NUM_GIFTS_4YRS $\geq$ 19, TRAVIS_RESIDENT = 1, and EX_4YRS = 0. The first two conditions are the same as for the previously studied pattern, while the last states that the members of this group have not given to the Texas Exes in 2015-2018. Bob finds this either-or relationship rather interesting.

This group is somewhat smaller than the previous group Bob had studied – it has 41 members – but the probability of its members to donate to the campus radio station is quite high – 75.6%. The additional condition EX_4YRS = 0 added another 11.4% of likelihood. Bob finds that he gathered very valuable information in this exploration. The conditions he found are not overly narrow or artificial. They make much sense to him and he decides to pass his findings over to the radio station to help them in fundraising drives.

Bob's journey is in fact a good example for how a user would use our system to continuously refine the characteristics of a certain family of groups and come up with nuanced multi-level marketing strategies. The marketer could first launch a more general campaign for a broader group (e.g. the group of section 4.3.1) and then address smaller but more specific groups with more targeted campaigns (e.g. the groups identified in sections 4.3.2 and 4.3.3) with higher probabilities of success.

5 Discussion

The case studies have provided some insight into how an analyst would conduct exploratory studies with our visual interface. The interface excels because it allows users to explore a dataset from multiple perspectives all within a single session from one dashboard. Analysts can quickly follow their instincts via simple mouse-click interactions and see the results not just as a single number but also with visual explanations of how the number was derived and how it relates to the overall data. As such our system fully embraces the paradigm of explainable machine learning and AI – it endows analysts with confidence that a recommendation is firmly grounded in reality.

A domain like advancement is a complex ecosystem where the many features can play different roles in shaping groups of donors, often in surprising and unexpected ways. While our case studies have only given a glimpse of this diverse ecosystem of patterns, each of these patterns was prescriptive in a sense that it gave clear characterizations on the particular type of donor that would be responsive to the donation type of interest. At the same time, the characterizations were sufficiently succinct and focused on the important features only. In that sense the explanations were minimally complete which is important for fundraising campaigns as they seek to spread the net as wide as possible.

Our general system embraces the fact that large populations are often decomposed into a set of homogenous subgroups. This is well known in fields like medicine and can be overcome by careful subgroup analysis, i.e. identifying the specific patient characteristics that benefit a desired outcome. Typically these selective features are determined either by prior knowledge, pre-specification, or a stepwise procedure, none of which is scalable in the number of features. In contrast, we learn these subgroups by discovery using fully automated statistically robust pattern mining which can scale to 1,000s and more features/variables.

Compared to linear/logistic regression models, our system has several distinct advantages as it becomes intractable to explicitly model all possible interactions between variables. Even if we only restrict ourselves to pairwise interactions the dataset we studied in this paper would have over 10,000 possible interactions. In addition, regression models are restricted to modeling linear relationships. Nonlinear relationships would require additional transformations to be captured. This is particularly labor intensive for higher dimensional datasets. In contrast, our system is able to identify interactions and capture nonlinear relationships automatically.

A key advantage of our system over black box models (e.g. random forests, neural networks, etc.) is that our system is designed for the human in the loop. Although explainable AI tools exist (such as SHAP, LIME, etc.) to help explain a black box model's decision, there is no guarantee that the model is basing this decision on a true cause-effect relationship or some spurious correlation. Conversely, our system displays alternative explanations (i.e. via the group summary plot) which allow the analyst to identify the most likely explanation for why a group is more likely to donate.

Finally our system is not dedicated to advancement only. It is very general in the spectrum of application areas in which it can be deployed. We have used it in areas as diverse as public health and epidemiology, fintech and finance, biotech and bioinformatics, computer systems analysis and configuration, and many more.

6 Conclusions

We have demonstrated that automated pattern analysis can be highly effective in defining the characteristics of donors more likely to make philanthropic contributions to a university. The patterns were extracted without any manual tuning of parameters. The visualizations are also automatically produced by our method and are helpful to understand the statistics that underlie these patterns and make the findings more accountable.

Acknowledgements

The system used for this analysis is marketed as a software package called the Pattern Browser developed by Akai Kaeru LLC (https://akaikaeru.com). Its development has been funded by the US National Science Foundation (NSF) via SBIR grant 1926949. We very much thank John Gough from U Texas, Austin for providing the data and his insight interpreting them.

References

[1] H. Kriegel, P. Kröger, A. Zimek, "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering," *ACM Trans. Knowledge Discovery from Data*, 3(1):1, 2009

[2] B. Wang, K. Mueller, "Does 3D really make sense for visual cluster analysis? Yes!" *IEEE VIS Workshop on 3DVis: Does 3D Really Make Sense for Data Visualization?* Paris, France, November 2014.

[3] P. McKnight, J. Najab, "Mann-Whitney U Test," *The Corsini Encyclopedia of Psychology*, 2010.

[4] J. Han, J. Pei, Y. Yin, "Mining frequent patterns without candidate generation," *ACM SIGMOD*, 29(2): 1:12, 2000.

[5] S. Lundberg, S. Lee, "A unified approach to interpreting model predictions," arXiv preprint arXiv:1705.07874, 2017.

[6] Four Steps to Identify Your Planned Giving Prospects https://resources.freewill.com/planned-giving-prospects