Predicting Customer Lifetime Value with Unified Customer Data

DATA SCIENCE GROUP, AMPERITY INC.

Customer lifetime value (CLV) modeling is the lynchpin of modern marketing analytics, allowing marketers to prioritize customers that have the highest predicted business value. Much of the literature (and much debate amongst industry practitioners) has focused on the difference in performance across different types of CLV models. But lost in this debate is the reality that modern retail businesses collect and have access to more data from channels than ever before: location information, itemized products, and product categories, as well as customer demographic information. Leavaraging this richer source of customer data should improve predictive accuracy: for instance, customers whose first purchase is of a certain product category may be more likely to churn, and this information should be available to a predictive attrition model. In this paper, we present an approach to CLV and churn prediction which uses this information via Amperity's unified customer data profile. We empirically compare our model to a state-of-the-art baseline that utilizes only historical transaction data (Extended Pareto/NBD) and demonstrate significant gains from leveraging unified customer data features across several retailers, forecasting horizons, and various prediction settings; our unified data modeling approach enables us to outperform this baseline by an average of 13.4% F1 in churn prediction and 15.2% improvement in CLV prediction RMSE error.

1 INTRODUCTION

A key foundation of customer-centric marketing is the ability to use data to identify which customers will have the highest business value [6]. The ability to do this accurately at scale allows marketing teams to focus their acquisition and retention efforts on high-value customer segments, while focusing less on customer segments that are unlikely to yield long-term value. The problem of identifying high-value customers has typically been formulated in terms of predicting customer lifetime value (CLV) given past customer behavioral data [11], and more specifically *order-level* transactional data. The seminal model in this area, the joint Pareto/negative binomial distribution (Pareto/NBD) model introduced in [14], operates exclusively on the time-series of transaction dates and total order values; effectively, this model reduces data for each customer to a few simple summary statistics, such as the time elapsed since the last transaction or the average of past transactions. While there are many popular variants of this model [7, 8], these variants do not change this core assumption or dramatically grow the scope of what kind of data is used for training or predicting CLV. In fact, a simple extension of the Pareto/NBD approach [8], remains state-of-the-art in a recent empirical comparison [12] across several retail datasets and published models.

While transactional data is clearly relevant to predictive CLV and retention, it is far from a complete picture of a customer or how they interact with a business. For instance, a customer being enrolled in a loyalty program or living in an area where a retail brand has strong word of mouth will impact the likelihood of that customer's retention [16], but these signals are ignored by traditional CLV approaches. Modern consumer businesses collect a wealth of information about their customers and interactions they have with a brand. One of the largest barriers to effectively utilizing this data to help drive better decision-making is the difficulty in unifying this data across business departments and channels. For this reason, there has been much interest in Customer Data Platforms (CDPs) which unify this data. The core aim of this work is to develop a modeling approach that maximizes the value from incorporating features derived from rich customer profiles that a CDP provides.

Specifically, in this work we'll describe Amperity's CLV prediction model, which is built on its unified Customer Data Platform, powered by Amperity Fusion™, which uses a machine learning approach to customer data unification [2, 3]. For reference, Amperity serves a variety of businesses, including retail, travel, hospitality, finance, insurance, and healthcare. The types of data available for each business vary, some with canonical itemized transactions (individual purchased items) and others with richer store location information. Although the results will only reflect predictive performance across a sample of three retailers, the design and details reflect the flexibility needed to be effective for the range of Amperity partners. The key contributions of this work are:

- An empirical exploration of which kinds of customer signals yield the most incremental improvements in CLV and churn modeling. The most immediate take-away is that non-transactional features can have a strong impact on model performance, increasing the benefits of having a unified customer data model. The kinds of signals that matter vary from business to business, suggesting the need for automated feature exploration for black-box deployment of CLV modeling for business data.
- A simple discriminative learning approach to CLV and churn prediction, which can more easily incorporate features using the broad range of available customer data. While our modeling approach is a simple combination of linear and logistic regression, it performs surprisingly well and can be used effectively over the use-cases Amperity faces, including making predictions over hundreds of millions of customers.
- A broad evaluation of CLV modeling across various predictive settings, including multiple businesses, forecasting horizons (quarterly and yearly), as well as two different prediction conditions (a post-purchase "triggered" condition, mirroring a common marketing use case, and a condition just based on a fixed-date, like most of the traditional CLV modeling settings). These settings reflect the varied use cases of predictive analytics in practice.

In Section 2 we will review the formal task setting for CLV and churn prediction, as well as the data setting we use for experiments and how we will evaluate approaches. In Section 3 we will provide an overview of CLV and churn prediction modeling approaches for our state-of-the-art baseline, as well as our approach and the unified customer data features we explore. In Section 4 we do a deep dive into understanding the performance our model (under various feature settings) and against the baseline approach.

2 SYSTEM OVERVIEW

In this section, we define the problem setting and the main objective of this paper, and provide an overview of our data architecture and the evaluation process.

2.1 Problem Statement

Customer lifetime value (CLV) and churn prediction are two of the most important metrics in customer-centric marketing analytics. CLV is the the total amount of money a customer will spend during their lifetime (as a customer of the business); it is comprised of a historical amount we already know the customer has spent and a future, or *residual*, lifetime spend that we are attempting to predict. For the remainder of this work, we will use CLV to refer to the residual lifetime value. Since the full lifetime is unknown, marketers will typically use a prediction for future CLV spend over a given forecasting horizon (e.g, the next quarter or year). These CLV forecasting metrics not only help the business make decisions about how much money to invest in acquiring new customers and retaining existing ones, they also enable the business to identify their most valuable customers.

A related problem to CLV prediction is churn prediction. A churned customer is one that has no transactions in a forecasting horizon (e.g, do not return in a quarter or year). As with CLV prediction, the systems described here will predict if a customer will churn in a given forecasting horizon. Businesses can use churn probabilities to help prioritize which customers to attempt to re-engage and prevent from churning or better understand what experiences may have led to their churn.

The last few years have seen an explosion of interest in Customer Data Platforms (CDPs) to help businesses unify customer data across IT systems, brands, and channels into a unified customer profile [9]. A primary contribution of this work is understanding how the rich machine learning features derived from a unified customer dataset can benefit predictive analytics modeling such as CLV and churn prediction.

We compare our feature-rich model to the Extended Pareto/NBD (EP/NBD) model [7], since a recent empirical "bakeoff" [12] demonstrated its performance was no worse than a broad range of CLV models. We specifically used the open-source Python 'Lifetimes' package.[1]

Like most of the literature in this area, the EP/NBD primarily focuses on RFM statistics (i.e., the basic transaction information such as recency, frequency, and monetary values). This serves as a good contrast for the relative value from richer customer profile features.

Our evaluation focuses on two predicted statistics commonly used in marketing segmentation:

- F_1 of churn prediction: The precision, recall, and F_1 of a prediction that a customer will return in a forecasting horizon (a quarter or a year here).
- RMSE of CLV prediction: The root mean-squared-error of CLV predictions for a given interval (future quarter or year) given customer data and past interactions. This is the most commonly used metric to understand predicted CLV accuracy.

We also generate examples for evaluating the above in two different scenarios:

- FIXED-DATE: Pick a universal evaluation cut-off date and train models with data up that date, then evaluate performance in the interval after that date. This condition is most common in the literature.
- POST-PURCHASE: Since marketing decisions typically happen following a recent purchase, we also focus on the setting of evaluating after each customer purchase, since this better reflects some application settings. This task is generally harder than the FIXED-DATE setting since you don't allow time to elapse from the most recent purchase.

2.2 Data Architecture

The high-level framework of our predictive CLV pipeline is shown in Figure 1. Raw customer data from various data sources is preprocessed and unified in our identity resolution pipeline. The identity of each record is resolved in the pipeline, and records that represent the same customer are merged together to create a unified customer profile. The final unified customer dataset is the collection of all unified customer profiles. It includes a variety of information such as customer attributes, transaction history, and purchased product attributes (see details in Section 2.3).

[1] https://github.com/CamDavidsonPilon/lifetimes

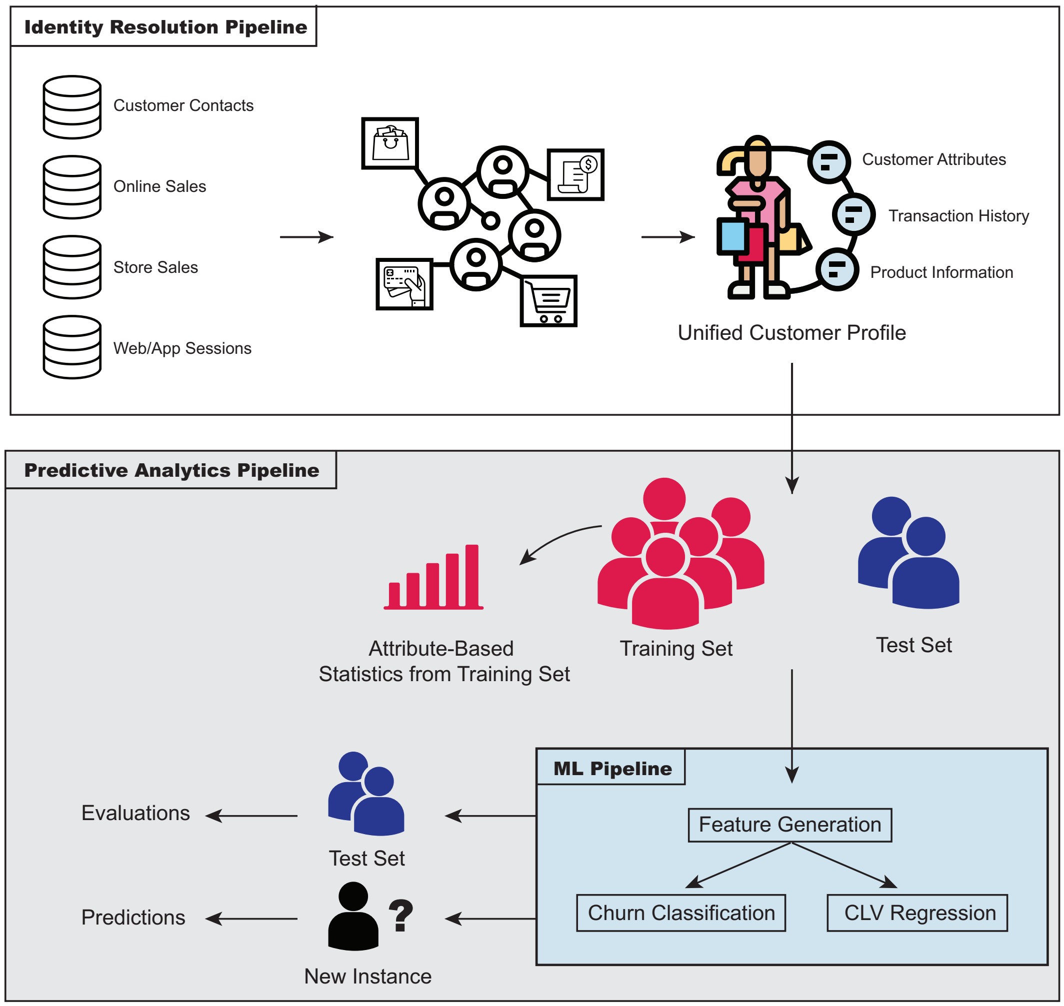

Fig. 1. High-level overview of the predictive CLV framework.

All the data are sampled from the unified customer dataset (with a 70/30 train/test split). One of the major challenges of predicting CLV is the unusual distribution of the target variable. Of the customers with greater than zero CLV, the values differ by several orders of magnitude. To make the evaluation process more meaningful and the training more effective, we perform stratified sampling based on CLV percentiles, and the customers with CLV of zero (customer with no following purchases within the horizon) occupy a single strata.

We also compute aggregated statistics for each attribute to support the later feature generation. Examples of attribute statistics include the observed average purchase frequency based on zip codes, and the probability of customers thhat purchase in the next forecasting interval given their previously purchased product IDs. This allows us to assess the impact of each categorical attribute's values such as names, zip codes, and product categories, and assign each attribute with numerical feature values (e.g., probability of purchasing in the next quarter when the customer's given name is "Jennifer"). The separation of generating the attribute-based statistics from the ML model also allows us to update these numerical feature values without retraining the model. We note that to prevent information leakage, the attribute statistics are generated from the training set only. Finally, the sampled data and the aggregated attribute statistics flow through an automated feature generator to generate customer features. We use the same set of features to train both churn classification and CLV regression models; we expect the most relevant features will differ between these tasks, but rely on regularization to help prune irrelevant signals. The performance of the models are evaluated on a separate test set.

2.3 Unified Customer Data

All data used in this work, both by our model and the baseline, is taken from three retailers on Amperity's Customer Data Platform; we anonymize these businesses as Retailer A, Retailer B, and Retailer C. [2]

Customer data has been unified using Amperity's identity resolution algorithm[2], which consolidates business data associated with the same real-world customer. A complete customer profile can include, but is not limited to a customer's demographic information, their entire purchase history with the business (such as purchase/return history), marketing touchpoints, web/app session logs, etc. In this paper, we show that rich features derived from a unified customer attributes and simple ML models yield significant wins in predicting churn and CLV.

The set of customer attributes used across datasets varies but typically includes personal information such as name, phone, email, birthdate, and address, as well as brand-related attributes such as preferred store ID, preferred shopping channel, top product category, etc. The order-level transaction history usually contains the high-level summary of all past orders, such as order date, order channel, and order total value. Among the three datasets, only Retailer A has complete itemized transaction data which we can use to link to the full product attributes. Retailers B and C have similar data, but it is for online orders only. The fields are broken down into three categories, "Customer Attributes", "Transaction History", and "Product Attributes". The data fields available to each dataset (and used in the CLV and churn modeling) are shown in Table 1.

2.4 Prediction Settings

The prediction settings we used in the CLV and churn modeling include 1) prediction mode and 2) prediction interval. In traditional CLV prediction, the starting date of a given prediction period is usually the same for all customers (i.e., FIXED-DATE mode). It is useful when the goal is to build scheduled marketing campaigns or to evaluate the business on a regular basis. However, to create a personalized experience and to add meaningful customer touch-points in time, being able to predict future purchase behavior right after each purchase event (i.e., POST-PURCHASE mode) is invaluable. In this paper, to the author's best knowledge, for the first time, CLV churn and prediction are modeled in both FIXED-DATE and POST-PURCHASE modes. The illustration of how the features and labels are generated is shown in Figure 2. Our approach (as with most ML models) requires fitting different models for each forecasting horizon (Δt in Figure 2). In this paper, the two most common prediction periods, long (52 weeks) and short (13 weeks), are evaluated to demonstrate the reliability of our approach.

[2] The datasets in Retailers A, B, and C contain 6.8 MM, 2.6 MM, and 11.2 MM customer records, respectively.

Data Class	Data Field	Dataset
Customer Attributes	first name	A, B, C
	last name	A, B, C
	birthdate	A, B, C
	city	A, B, C
	state	A, B, C
	zip code	A, B, C
	preferred store ID	A
	nearest store ID	A
	nearest store distance	A
	preferred product category	A
	email subscriber?	B
	email open rate	B
	friends or family?	B
	loyalty accounts	C
Transaction History	order ID	A, B, C
	order date	A, B, C
	order channel	A, C
	order total value	A, B, C
Product Attributes	product ID	A, B, C
	product name	A, B, C
	product category	A, B, C
	product quantity	A, B, C

Table 1. The data fields used in the CLV modeling in each retailer's dataset.

3 PREDICTIVE MODELING

CLV prediction is a regression task to predict the total monetary spend of a given customer over the next Δt days using historical interaction information with the business. A related task is churn prediction, predicting whether or not the customer will return at all (i.e, have *any* spend) over the next Δt days. The focus of this work is to show improvements on both of these by using richer customer-centric features derived from unified customer data (e.g, demographics, proximity to store location), as well as richer information about past interactions (e.g, specific products purchased).

In order to more easily utilize rich feature representations, we chose to move away from the generative probabilistic models [7, 8] that are dominant in CLV and churn prediction. Before describing our approach, we will review the commonly used generative probabilistic models, the highest-performing of which we use as a baseline.

3.1 Baseline Approach

As a baseline we adopt the Extended Pareto/NBD model (EP/NBD) described in [7]; this model has been shown to have state-of-the-art performance according to a recent empirical evaluation [12] of CLV prediction techniques.

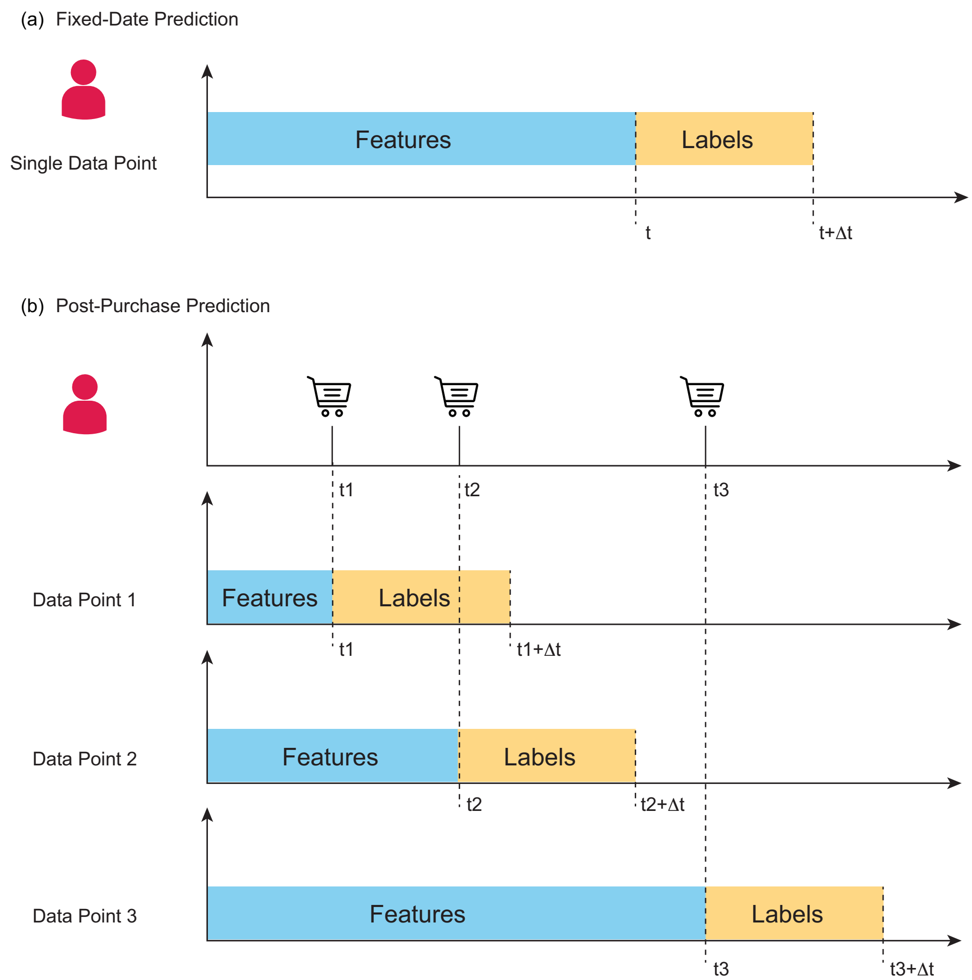

Fig. 2. Illustration of how features and labels are generated for a single customer using two different prediction modes, (a) the FIXED-DATE prediction mode (b) the POST-PURCHASE prediction mode.

The EP/NBD model is a generative probabilistic model which generates the time-series of transaction time and order amounts. The model makes the following assumptions [15] in order to first generate a sequence of purchase times:

- Purchase count follows a Poisson distribution with the rate λ.
- Lifetime distribution follows an exponential distribution with the slope μ.
- The latent parameters λ and μ are constrained by two prior gamma distributions representing how these latent parameters are distributed among the population of customers.

The Pareto/NBD only focuses on purchase counts and lifetimes, and it does not generate the monetary value of each purchase. Typically the Pareto/NBD model is coupled with a modeling choice for generating order values; the EP/NBD model utilizes the Gamma-Gamma extension to the Pareto/NBD model.

The Gamma-Gamma component makes a few more assumptions [10]:

- At the customer level, the transaction/order values vary randomly around each customer's average transaction value.
- The observed mean value is an imperfect metric of the latent mean transaction value $\mathbf{E}(M)$, where M represents the monetary value.
- Average transaction value varies across customers, though these values are stationary. This is a big assumption and may not be valid in many business situations.
- The distribution of average values across customers is independent of the transaction process. In other words, monetary value can be modeled separately from the purchase count and lifetime components of the model. This may or may not hold in typical business situations.

The Pareto/NBD model computes the expected number of purchases in the next prediction interval for each customer and the Gamma-Gamma model assigns a value to each of those future purchases. Together, the CLV for each customer in the next prediction interval can be predicted. As the most popular CLV modeling technique, the Extended Pareto/NBD model has several advantages:

- Once trained, it can be used for any time window. You don't need different models for the short-term and the long-term forecasting.
- The data requirements are simple: you only need to collect RFM values.
- The model only has seven parameters (four for Pareto/NBD and three for the Gamma-Gamma extension).

However, this model also has several disadvantages:

(1) It is challenging to add arbitrary new features.
(2) The assumptions of the generative model may hold in aggregate but are not flexibile enough for individual customer behavior to differ.
(3) Strong assumptions, such as stationary average transaction values over time may not hold for many businesses.
(4) The model is weaker for customers who have a short transaction history, and is not applicable customers without transaction history.

3.2 Our Approach

Rather than describe a probabilistic process for generating the time-series of transactions, we instead cast CLV and churn prediction as discriminative ML problems. Specifically, using churn prediction, we utilize a logistic regression model to predict the binary event of whether a given customer will return over the next Δt days given past interactions and customer information. Similarly, we cast CLV prediction as a linear regression problem that predicts total spend over a horizon of Δt days conditioned on customer information and past interactions.

One significant detail to note is that each of these models is trained for a single value of Δt days as a forecasting horizon; in contrast, the EP/NBD model can flexibly answer questions about *any* horizon. In practice, this isn't a strong disadvantage since businesses normally require only a small number of time horizons and model can be trained for each horizon (e.g, monthly, quarterly, yearly). Given that the complexity of each model is relatively small, this poses no great computational burden.

Many different learning algorithms have been proposed for CLV and churn prediction: such as Markov chains and decision trees in [13], random forests in [4], and deep neural networks in [5]. The primary criterion driving our choice of logistic and linear regression was the relative simplicity of these model classes and being able to easily experiment with feature engineering on unified customer data, while maintaining high-performance at scale.

We explore three different approaches for end-to-end CLV prediction which compose linear models:

- Approach 1: use linear regression to model CLV directly, i.e.,
$$\text{CLV} = \mathbf{w}^T \ \mathbf{x} + \mathbf{b} \,, \tag{1}$$
Where $\mathbf{x}$ is a feature vector and $\mathbf{w}$ and $\mathbf{b}$ are the feature coefficients and intercepts, respectively.
- Approach 2: Train separate models to predict whether a customer returns in a time horizon (churn) and then conditioned on the return, predict the CLV spend in the horizon using a linear regression model:
$$\text{CLV} = P(\text{return?}|\mathbf{x}) \cdot \left(\mathbf{w}^T \ \mathbf{x}_{\text{returned}} + \mathbf{b}\right) \tag{2}$$
where $P(\text{return?}|\mathbf{x})$ comes from a logistic regression model predicting whether the customer churns in the forecasting horizon. Note that the CLV regression model trained in this approach is trained only on examples where the customer returns in the prediciton horizon. This also corresponds to the expected spend assuming we independently decide if a customer returns and then generate spend using a linear regression model.
- Approach 3: Rather than use a linear regression model to predict CLV as in Approach 2, instead predict the number of transactions in the prediciton horizing using linear regression, and then multiply by empirical average order value to estimate CLV spend. As in Approach 2, utilize a logistic regression model to predict the binary return event:
$$\text{CLV} = P(\text{return?}|\mathbf{x}) \cdot \hat{\nu}_{\text{returned}} \cdot \hat{\zeta}_{\text{returned}} \,, \tag{3}$$
Where $\hat{\nu}_{\text{returned}}$ is the predicted purchase count (from a linear regression model) conditioned on return, and ζ_{returned} is the hisotircal average purchase value for the returned customer.

The performance of the three approaches described above varies with different prediction intervals (Δt) and prediction settings (described in Section 2.4). For each setting, we report prediction quality for the approach that yields the lowest RMSE compared using the training data (a similar hyper-parameter search is done for the baseline).

3.3 Features

With unified customer datasets, our models can incorporate features across several kinds of customer informatio. There are four broad classes of features we explored:

(1) **Customer Attributes**: We define features for customer attributes using statistics about the behavior of customers with the attribute. For example, we may have a feature for the average number of orders or cart total for customers from a given city. Similar feature variants exist for most discrete demographic attributes of

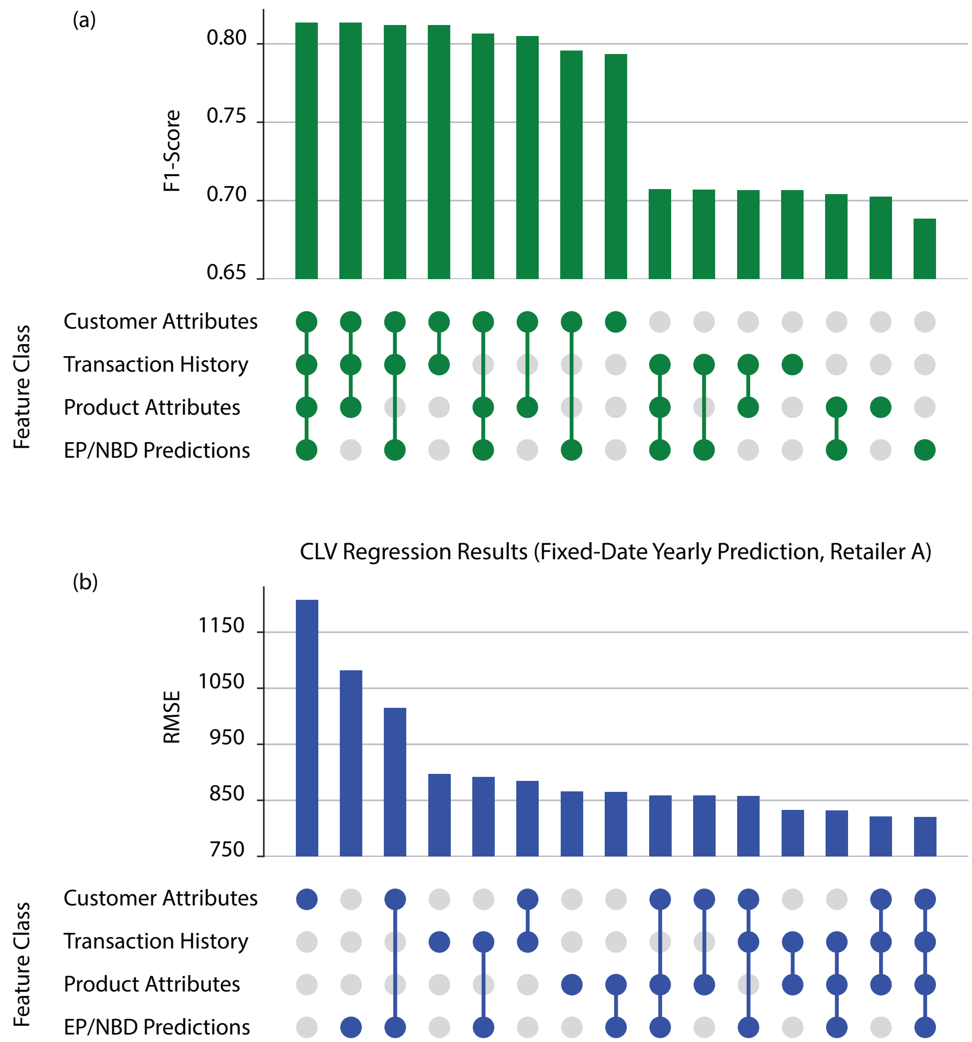

Fig. 3. An example of the feature ablation study performed on Retailer A for $\Delta t = 365$ days and in FIXED-DATA mode. Every combination of the four feature groups described in Section 3.3 are evaluated against churn prediction F_1 (higher is better) and CLV prediction RMSE (lower is better). Note the general lift in churn prediction coming from "Customer Attributes", validating the hypothesis that unified customer data can help improve churn risk modeling.

a user; note that the statistics never utilize customer data in the test set so these statistics only reflect training customer historical data.

(2) **Transaction History**: These are variants of RFM statistics (e.g, average order value) computed over various historical windows (last quarter, year, etc.). We may also use features like the month of the most recent transaction to capture seasonality information.
(3) **Product Attributes**: Attributes about specific cart items when available, mainly derived from looking at the "stickiness" of product purchases, defined by the likelihood of a return purchase within Δt days after

Rank	Feature Name	Class
1	Order frequency predicted by EP/NBD model	(4)
2	Number of distinct products purchased in the past	(3)
3	Probability of a return purchase in the next year given preferred product category	(1)
4	Order frequency in the past half year	(2)
5	Probability of a return purchase in the next year given preferred store id	(1)
6	Aged from 40 to 50	(1)
7	Aged from 50 to 60	(1)
8	Average past order frequency given the last 5 purchased products	(3)
9	Probability of a return purchase in the next month given preferred product category	(1)
10	Aged from 30 to 40	(1)

Table 2. The top 10 most important features of the yearly churn classification model in the POST-PURCHASE mode for Retailer A, as well as the four different feature classes they belong to, i.e., (1) Customer Attributes, (2) Transactional History, (3) Product Attributes, (4) EP/NBD Predictions. The absolute value of model weight coefficients is used as measure of feature importance.

purchasing the item. We use this statistic defined at the level of an individual product to define features for a customer based on the average stickiness of all past product purchases as well as the maximum stickiness.

(4) **EP/NBD Predictions**: In order to understand if the EP/NBD captures any more information about the problem, we also experiment with adding its predictions (for frequency and average order value) as features to the models.

We perform a feature ablation study to understand how the model performance is impacted by the availability of the above feature classes. Figure 3 shows an example of the fixed-date yearly prediction for Retailer A, who has the richest data sources for features among the three retailers. In Table 2, we also give an example of the top 10 most important features of the yearly churn classification model in the POST-PURCHASE mode (trained for Retailer A). It is expected that with all four feature classes incorporated, both the churn classification and the CLV regression perform the best (i.e., highest F_1-score in the churn classification and lowest RSME in the CLV regression).

Notably, the EP/NBD prediction features add little marginal value with the other feature sets present, suggesting the information captured by this model is redundant with other feature classes,despite having high feature importance. We hypothesize this is because the other transaction features in the model capture similar trends, but dispersed across many features. It is somewhat suprirsing that by using customer attributes alone, we can achieve predictions almost as accurate as when using all four feature classes, with the F_1 score of 0.79 vs 0.82. This indicates that with rich customer attributes (mostly just demographic information) the model has the capability to predict churn without the need for any transaction history. This is significant because most statistical models, including the Pareto/NBD model, cannot make inferences on customers who have no past purchase history. Another surprise is that by only using only features generated from previously purchased products,(Product Attributes group above), we are able to predict the CLV almost as accurately as by using all feature classes, while in the traditional EP/NBD setting, this also couldn't be incorporated.

F1 of Churn Classification	Quarterly (90 days)			Yearly (365 days)		
Dataset	EP/NBD	OURS	$\Delta(F_1)$	EP/NBD	OURS	$\Delta(F_1)$
Retailer-A	0.52	0.62	+19.52%	0.69	0.82	+18.93%
Retailer-B	0.46	0.55	+19.84%	0.59	0.79	+33.23%
Retailer-C	0.63	0.72	+14.05%	0.70	0.80	+13.58%
Mean	0.53	0.63	+17.47%	0.66	0.80	+21.91%

Table 3. Churn prediction results in the FIXED-DATE mode. AMP is the Amperity system described in this work (see Section 3.2) and EP/NBD is the baseline described in Section 3.1.

4 RESULTS AND ANALYSIS

In this section, we present the results for both churn classification and CLV regression in four different prediction settings, parameterized by the prediction window (Δt) and the prediction mode (m), where $(\Delta t, m) \in$ {(365 days, POST-PURCHASE), (90 days, POST-PURCHASE), (365 days, FIXED-DATE), (90 days, FIXED-DATE)} , using three different retailers' datasets. We show that, compared to the baseline model, our approach, labeled AMP, improves the F_1 score in churn classification by 19.68% and 7.11% on average in fixed-date and post-purchase modes, respectively; and reduces the RSME error in CLV regression by 20.76% and 9.61% in these two modes, respectively. The results in each individual setting is shown in Tables 3 to 6. Recall as described in Section 3.2, these results are obtained from the model with the highstest training score.

Compared to the EP/NBD model, across all settings, our approach yields a meaningful F_1-score lift in churn classification and RSME reduction in CLV regression.We observe that both models perform best when using a longer prediction interval and predicting from a fixed date. It is not surprising, and the possible explanations are 1) when the prediction interval is long, the customer is more likely to have enough time to establish a predictable purchase pattern; 2) when forecasting starting from a fixed date, the model uses the customer's entire purchase history before that date, which may contain multiple purchases, and therefore, the information contained in the features is richer. Higher quality features should help build a better model. It is worth noting that the improvement from our model tends to be larger in this setting.

4.1 Future Work

When the prediction interval is short or when we try to predict the future after a single purchase event, the prediction quality degrades as expected. Although we outperform the baseline EP/NBD model across all settings, we believe there is still significant room for improvement. For example, we have not incorporated clickstream data or customer touch-points (e.g, email campaigns) into our model, and these types of interactions usually play a larger role in short-term or event-triggered customer activities. Recent work in the search ranking literature [1] have successfully leveraged interaction data such as this to improve search ranking personalziation and similar techniques might benefit predictive modeling.

RMSE of CLV	Quarterly (90 days)			Yearly (365 days)		
Dataset	EP/NBD	OURS	Δ(RSME)	EP/NBD	OURS	Δ(RSME)
Retailer-A	567.75	523.52	-15.59%	1101.55	808.85	-26.32%
Retailer-B	87.10	82.72	-5.03%	209.51	175.73	-16.12%
Retailer-C	226.43	197.49	-12.78%	597.72	306.12	-48.79%
Mean	293.76	267.91	-11.13%	636.26	430.23	-30.40%

Table 4. RMSE for CLV prediction in the FIXED-DATE mode. AMP is the Amperity system described in this work (see Section 3.2) and EP/NBD is the baseline described in Section 3.1.

F_1 of Churn Classification	Quarterly (90 days)			Yearly (365 days)		
Dataset	EP/NBD	OURS	Δ(F1)	EP/NBD	OURS	Δ(F_1)
Retailer-A	0.51	0.57	+11.7%	0.67	0.76	+13.4%
Retailer-B	0.44	0.46	+3.39%	0.58	0.64	+11.7%
Retailer-C	0.75	0.76	+ 1.33%	0.88	0.89	+1.12%
Mean	0.56	0.60	+5.47%	0.71	0.76	+8.74%

Table 5. Churn prediction results in the POST-PURCHASE mode. AMP is the Amperity system described in this work (see Section 3.2) and EP/NBD is the baseline described in Section 3.1.

RMSE of CLV	Quarterly (90 days)			Yearly (365 days)		
Dataset	EP/NBD	AMP	Δ(RSME)	EP/NBD	AMP	Δ(RSME)
Retailer-A	437.53	384.64	-13.75%	1296.68	1140.27	-12.06%
Retailer-B	95.13	82.89	-9.85%	286.52	250.01	-12.74%
Retailer-C	231.77	220.78	-4.74%	555.86	530.68	-4.53%
Mean	254.81	229.44	-9.44%	713.02	652.48	-9.77%

Table 6. The CLV prediction results in the POST-PURCHASE mode. AMP is the Amperity system described in this work (see Section 3.2) and EP/NBD is the baseline described in Section 3.1.

5 CONCLUSIONS

Modern retail business collect more data from more channels than ever before, but traditional approaches to CLV and churn prediction only use a limited subset of data per customer, typically just the time-series of their transaction frequency and order values. Amperity's approach to CLV modeling utilizes a much more comprehensive customer dataset unified by Amperity's Customer Data Platform. In this paper, we demonstrated how the rich machine learning features derived from the unified customer dataset benefits predictive analytics modeling such as CLV and churn prediction. We empirically compared our model to a state-of-the-art baseline (Extended Pareto/NBD) and demonstrated significant improvements in three different retailers' datasets and under various prediction settings. Our model achieved on average 13.4% F1 improvement in churn classification and 15.19% RSME reduction in CLV prediction.

REFERENCES

[1] Qingyao Ai, Yongfeng Zhang, Keping Bi, Xu Chen, and W. Bruce Croft. Learning a hierarchical embedding model for personalized product search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 645–654, New York, NY, USA, 2017. ACM.

[2] Amperity. Amperity intelligent identity resolution. https://amperity.com/resources/whitepaper/intelligent-identity-resolution, 2018.

[3] Amperity. How amperity brings all your rich customer data together. https://amperity.com/resources/whitepaper/how-amperity-brings-all-your-rich-customer-data-together, 2018.

[4] Benjamin Paul Chamberlain, Angelo Cardoso, Chak H Liu, Roberto Pagliari, and Marc Peter Deisenroth. Customer lifetime value prediction using embeddings. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1753–1762. ACM, 2017.

[5] Pei Pei Chen, Anna Guitart, Ana Fernández del Río, and África Periáñez. Customer lifetime value in video games using deep learning and parametric models. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2134–2140. IEEE, 2018.

[6] P. Fader and S.E. Toms. *The Customer Centricity Playbook: Implement a Winning Strategy Driven by Customer Lifetime Value*. Wharton Digital Press, 2018.

[7] Peter Fader, Bruce Hardie, and Ka Lee. "counting your customers" the easy way: An alternative to the pareto/nbd model. *Marketing Science*, 24:275–284, 05 2005.

[8] Nicolas Glady, Bart Baesens, and Christophe Croux. A modified pareto/nbd approach for predicting customer lifetime value. *Expert Syst. Appl.*, 36(2):2062–2071, March 2009.

[9] Paul Greenberg. How customer data platforms can benefit your business. https://www.zdnet.com/article/introduction-to-customer-data-platforms/, Feb 2017.

[10] Sunil Gupta, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V Kumar, Nathaniel Lin, Nalini Ravishanker, and S Sriram. Modeling customer lifetime value. *Journal of service research*, 9(2):139–155, 2006.

[11] Dipak Jain and Siddharth Singh. Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, 16:34 – 46, 03 2002.

[12] Pavel Jašek, Lenka Vraná, Lucie Sperkova, Zdenek Smutny, and Marek Kobulsky. Modeling and application of customer lifetime value in online retail. *Informatics*, 5, 01 2018.

[13] Peter Paauwe, Peter van der Putten, and Michiel Van Wezel. Dtmc: An actionable e-customer lifetime value model based on markov chains and decision trees. *ICEC*, 7:19–22, 2007.

[14] David C. Schmittlein, Donald G. Morrison, and Richard Colombo. Counting your customers: Who are they and what will they do next? *Manage. Sci.*, 33(1):1–24, January 1987.

[15] David C Schmittlein, Donald G Morrison, and Richard Colombo. Counting your customers: Who-are they and what will they do next? *Management science*, 33(1):1–24, 1987.

[16] Jason Zhang, Ashutosh Dixit, and Roberto Friedmann. Customer loyalty and lifetime value: An empirical investigation of consumer packaged goods. *The Journal of Marketing Theory and Practice*, 18:127–140, 04 2010.