Customer Lifetime Value (CLV): Concepts and Modeling Approaches

1. What is Customer Lifetime Value (CLV)?

Customer Lifetime Value (CLV) is the expected net profit that a business earns from a customer over the entire duration of their relationship. Conceptually:

CLV = Expected discounted sum of future profits from a customer.

In general form, for customer i:

$$CLV_i = \Sigma_t [E(margin_i(t)) / (1 + r)^t],$$

where $margin_i(t)$ is the profit at time t and r is the discount rate.

2. ARPU (Average Revenue Per User)

ARPU stands for Average Revenue Per User (or per Unit) over a given period (typically month or year). It is widely used in SaaS, telecom, and subscription businesses.

If total revenue in a period is R and there are N active customers/users in that period:

$$ARPU = R / N$$

A simple steady-state CLV approximation is:

$$CLV \approx ARPU \times margin\% \times average\_customer\_lifetime$$

or, assuming a constant churn rate c per period:

$$CLV \approx ARPU \times margin\% / c$$

3. Historical (Descriptive) CLV

Historical or descriptive CLV measures how valuable a customer has been in the observed data window (e.g., 2009–2011). Using transaction data with order revenues $order\_revenue_{ij}$ for order j of customer i:

$$CLV\_hist_i = \Sigma_j order\_revenue_{ij}$$

We can also derive lifetime and average value:

$$lifetime\_days_i = last\_order\_date_i - first\_order\_date_i + 1$$
$$lifetime\_months_i \approx \max(lifetime\_days_i / 30, 1)$$
$$avg\_monthly\_value_i = CLV\_hist_i / lifetime\_months_i$$

This CLV is backward-looking only: it summarises past value, not future expectations.

4. Predictive CLV with Discrete Retention (Geometric Model)

Here we assume each customer i has a per-period retention probability p_retain,i and an expected margin per period $annual\_margin_i$. With discount rate r and time horizon T, a simple discrete-time CLV is:

$$CLV_i = \Sigma_{t=1}^T [annual\_margin_i \times p\_retain,i^t / (1 + r)^t]$$

The retention probability $p_{retain,i}$ typically comes from a churn model:

$$p_{churn,i} = P(\text{churn within horizon} \mid x_i)$$
$$p_{retain,i} = 1 - p_{churn,i}$$

Here x_i is the feature vector describing customer i (usage, tenure, prices, channel, segments, etc.). The churn model can be logistic regression, gradient boosting (e.g., XGBoost), or any other classifier. This is a discrete-time approximation with a constant retention probability per period.

5. Predictive CLV with Survival Models (Cox Proportional Hazards)

Survival analysis models the time until churn instead of a one-shot churn label. In Cox Proportional Hazards (Cox PH), the hazard (instantaneous risk of churn) for customer i at time t is:

$$h_i(t) = h_0(t) \times \exp(\beta^T x_i)$$

where $h_0(t)$ is the baseline hazard and x_i are features. From a Cox model we obtain the survival function:

$$S_i(t) = P(T_i > t \mid x_i)$$

which is the probability that customer i is still active at time t.

Then CLV can be approximated as:

$$CLV_i \approx \Sigma_{t=1}^T \left[E(margin_i(t)) \times S_i(t) / (1 + r)^t \right]$$

Survival models provide time-dependent retention curves and correctly handle censoring (customers still active at the end of the observation window).

6. RFM (Recency, Frequency, Monetary)

RFM is a classic framework to summarise customer behaviour using three dimensions:

- Recency (R): time since the last purchase.
- Frequency (F): number of purchases in the observation window.
- Monetary (M): total or average spend.

Use cases:

- Segment customers into bands (e.g., high-R/high-F/high-M as top VIPs, low-R/low-F as at-risk).
- Use RFM features as inputs to churn models or direct CLV regression.

RFM itself does not define a probabilistic CLV formula, but provides high-signal features strongly correlated with future value.

7. Non-contractual CLV: BG/NBD and Gamma–Gamma

For non-contractual businesses (e-commerce, retail, app usage), customers can become inactive without explicitly cancelling. A popular probabilistic approach is the BG/NBD

(Beta-Geometric / Negative Binomial Distribution) model for purchase frequency, often combined with a Gamma–Gamma model for monetary value.

High-level BG/NBD idea:
- Customers have different (unobserved) purchase rates and dropout probabilities.
- Purchase counts follow a Negative Binomial distribution driven by a latent Poisson rate.
- Dropout is modelled as a Geometric process with a latent probability.
- Across the population, these latent parameters are assumed to follow Gamma (for rates) and Beta (for dropout probabilities).

This structure can be viewed as a Bayesian hierarchical model with conjugate priors (Gamma and Beta), but in practice the model parameters are often estimated by maximum likelihood (frequentist). The 'Beta' and 'NBD' names refer to the marginal distributions that arise when integrating over these priors.

Given a customer's observed recency and frequency, BG/NBD provides the expected number of future transactions in a horizon. The Gamma–Gamma model is then used to model the distribution of average transaction value across customers. Combining them gives a probabilistic estimate of future revenue (CLV) for each customer.

8. Machine Learning CLV Models

Instead of closed-form models, we can learn CLV directly with machine learning.

8.1 Direct future value regression
- Define a target y_i = revenue or margin in the next K months from a reference time.
- Build features at the reference time (RFM, behaviour, prices, channels, cohorts, etc.).
- Train regressors (e.g., XGBoost, Random Forest, Neural Networks) to predict y_i.
- Optionally discount the horizon to obtain a net present value.

8.2 Two-stage frequency–monetary modelling
- Stage 1: model expected number of future orders.
- Stage 2: model expected margin per order.
- CLV_i ≈ E[#orders_i] × E[margin_per_order_i] (appropriately discounted).

These methods are flexible and capture non-linear effects, at the cost of less interpretability compared to classical models like BG/NBD.

9. Markov Chain / State-based CLV

In state-based CLV, customers move between discrete states (Active, Dormant, Churned, Bronze, Silver, Gold, etc.) over time. We estimate a transition matrix P where P_{ab} is the probability of moving from state a to state b in one period.

If each state s has an expected margin M(s), and X_t is the state at time t, then CLV can be written as:

$$CLV_i = \Sigma_{t=1}^T [E(M(X_t) | X_0, P) / (1 + r)^t]$$

This framework captures both churn risk and transitions between different value tiers.

10. Action-based CLV and Reinforcement Learning

In many applications, we want to know CLV under specific marketing or product strategies (e.g., discounts, emails, CRM actions). Here CLV becomes a function of a policy π that maps states to actions:

$$CLV_i(\pi) = E[\Sigma_{t=1}^T m_i(t, a_t) / (1 + r)^t | \pi]$$

where a_t is the action chosen by policy π at time t. We can use uplift modelling, contextual bandits, or full reinforcement learning to optimise π for long-term value, not just short-term response.

11. Profit-based (Cost-aware) CLV

CLV should ideally be defined in terms of profit, not just revenue. This means accounting for:

- Acquisition costs (CAC).
- Ongoing service/support costs.
- Marketing and discount costs.

A profit-based CLV might be written as:

$$CLV_i = \Sigma_{t=1}^T [E(revenue_i(t) - cost_i(t)) / (1 + r)^t] - CAC_i$$

This aligns CLV more directly with business value and decision-making about acquisition and retention investments.

12. Summary and Comparison of CLV Variants

Conceptually, all CLV approaches share the same idea: CLV is the expected discounted profit from a customer over time. The main differences between methods come from how we model:

- Purchase frequency and churn (simple churn rate, discrete models, survival/Cox PH, BG/NBD, ML).
- Monetary value (fixed ARPU, Gamma–Gamma, direct regression, cost-aware profit).
- Time (discrete periods vs continuous time survival).
- Actions and policies (static behaviour vs action-dependent CLV using RL or uplift).

A practical roadmap is:

1) Start with historical CLV and RFM segmentation.
2) Add churn models and simple geometric CLV (ARPU + retention).
3) For non-contractual settings, explore BG/NBD + Gamma–Gamma.

4) For richer setups or high stakes, use ML and survival models to refine CLV and support targeted retention, pricing, and acquisition strategies.