D. A. TSENOV ACADEMY OF ECONOMICS

Faculty of Management and Marketing

Department of Marketing

Prof. Todor Borisov Krastevich, PhD

CUSTOMER LIFETIME VALUE

(Conceptual, methodological, and applied aspects)

ABSTRACT

of a dissertation

for the acquiring of the academic degree "Doctor of Science"

Professional field 3.8 "Economics"

Scientific specialty "Marketing"

Svishtov

2026

The dissertation contains 318 pages and consists of an introduction, a presentation structured in five chapters, a synopsis, and a reflective epilogue. An integral part of the dissertation are two appendices containing a reproducible analytical protocol for modeling customer lifetime value, as well as a list of the 264 information and literature sources used. The exposition and appendices include 31 figures, 29 tables, and numerous online references to the data used and publicly available program code.

The dissertation has been discussed and referred for public defense by the Marketing Department at the Faculty of Management and Marketing of the D. A. Tsenov Academy of Economics in Svishtov.

Scientific jury:

Prof. Simeon Denev Zhelev, D.Sc.

Prof. Evgeni Petrov Stanimirov, PhD

Prof. Bistra Konstantinova Vasileva, PhD

Prof. Hristo Ivanov Katrandzhiev, PhD

Prof. Krasimir Marinov Marinov, PhD

Prof. Galina Pencheva Mladenova, PhD

Assoc. Prof. Ivan Rashkov Marchevski, PhD

The public defense of the dissertation will take place on May 29, 2026, at 01 p.m. in the "Rectorate" conference room of the D. A. Tsenov Academy of Economics – Svishtov. The materials related to the defense are available to interested parties at the Doctoral Studies and Academic Development Department of the D. A. Tsenov Academy of Economics – Svishtov.

Table of Contents

I. General Characteristics of the Dissertation

1.1. Relevance of the Research Topic

The dissertation argues the relevance of the issue of customer lifetime value (CLV) through three complementary perspectives. First, CLV is positioned as an economically sound bridge between strategic marketing and customer relationship management, as it allows decisions on customer acquisition, retention, and development to be subject to the logic of net present value and long-term return on marketing investment.

Second, the dissertation examines CLV in the context of widespread digitalization and customer-centric business models, where organizations have access to high-frequency and heterogeneous data, and the customer journey takes place in a multichannel environment. This changes the empirical basis and requirements for models, making analytical approaches capable of working with behavioral dynamics, nonlinearities, and interactions increasingly relevant.

Third, the study systematizes the impact of machine learning on CLV modeling, emphasizing the expansion of the scope of modeling, both in terms of the data used and the behavioral patterns captured, including in dynamic and "multimodal" customer environments.

From these standpoints, customer lifetime value is argued to be a key strategic tool for linking marketing decisions to the economic logic of value, especially in the context of digitalization, customer-centric business models, and the growing role of machine learning in marketing analytics. The emphasis is on the need for methodological integration of classical probabilistic approaches and modern predictive methods so that value forecasts are simultaneously accurate, interpretable by management, and operationally applicable.

1.2. Object, Subject, and Scope of the Study

The object of the study comprises the processes of evaluation, prediction, and use of customer lifetime value in marketing analytics and in the strategic management of customer relationships, with a focus on long-term customer–firm interactions, conceptualized as a value stream over the lifecycle of the relationship.

The subject of the study is the development, testing, and comparative analysis of models for predicting CLV, with a specific focus on machine learning algorithms. This includes the identification of predictors, model selection and training, comparison between classical probabilistic and contemporary machine learning approaches, as well as the evaluation of predictive accuracy, interpretability, and business applicability.

The scope of the research is defined functionally, methodologically, and empirically. Functionally, the dissertation treats CLV as a tool for marketing decision-making, including targeting, optimization of acquisition and retention costs, and the management of customer equity. Methodologically, the scope encompasses both classical statistical and probabilistic models, for example Pareto/NBD and Gamma-Gamma, and contemporary machine learning methods, for example Random Forest and XGBoost. Empirically, the analysis is based on simulated CRM-type customer databases, structured at the individual level, containing historical transactions and behavioral characteristics. The baseline context is e-commerce and services in a non-contractual setting, in which customer churn represents a latent process.

1.3. Research Objective, Tasks, and Working Hypotheses

The primary research objective is to assess the extent to which predictive machine learning approaches can enhance both the forecasting accuracy and the business usefulness of models for estimating and predicting Customer Lifetime Value (hereafter referred to as CLV), by developing and testing a hybrid methodological framework that combines probabilistic CLV models with machine learning algorithms, and by demonstrating its performance on simulated CRM-type data closely resembling real-world settings.

The operationalization of this objective is achieved through a set of sub-objectives and research tasks, including the derivation of managerial insights for integrating predicted CLV values into marketing decision-making processes and the formulation of best practices for implementing CLV analytics within organizations.

The research objective and tasks are articulated as a coherent framework aimed at: (1) systematizing the conceptual and methodological foundations of CLV, (2) developing a modeling and time-oriented validation protocol, (3) prototyping and comparatively testing alternative models, and (4) deriving actionable rules and managerial decision scenarios.

The working hypotheses are structured into six propositions (H1–H6) designed to combine theoretical grounding, empirical testability, and managerial relevance, based on the distinction between contractual versus non-contractual customer relationships and continuous versus discrete occurrence of transactional opportunities. The hypotheses address predictive accuracy in non-contractual and contractual settings, interpretability and transparency, cost efficiency and organizational feasibility, managerial applicability, as well as a hybrid thesis concerning the synthesis of the strengths of probabilistic and machine learning approaches. The hypotheses are formulated as follows:

H1: (Predictive accuracy in a non-contractual environment with irregular purchases): Machine learning models trained on large-scale behavioral and demographic datasets achieve higher predictive accuracy than probabilistic models when customer purchases are irregular and highly heterogeneous.

H2: (Calibration and interpretability): Probabilistic models provide better-calibrated predictions and higher interpretability compared to “black-box” models, making them more suitable under requirements for managerial accountability.

H3: (Contractual environment and churn risk): In contractual (subscription-based) relationships, models that explicitly predict churn risk, for example through survival analysis and/or classification algorithms, improve CLV estimation relative to simplified aggregated formulas.

H4: (Limited data availability): Under conditions of limited data, probabilistic models exhibit an advantage due to their stronger structural assumptions and greater stability of estimates.

H5: (Managerial usefulness): Model evaluation should also be validated through its impact on managerial decisions, such as targeting customer segments based on predicted CLV and the expected incremental return on investment or margin, rather than solely through statistical accuracy metrics..

H6: (Hybrid approach): The optimal practical solution is often a structured combination of behavioral theory, embodied in probabilistic models, and predictive flexibility, provided by machine learning, tested sequentially in non-contractual and contractual case settings.

1.4. Research Design and Limitations

The research design is comparative, integrative, and protocol-driven. At the theoretical level, two main classes of approaches to CLV forecasting are distinguished, probabilistic models for customer base analysis and predictive machine learning models, following the methodologically significant distinction between contractual and non-contractual customer relationships, as well as between continuous and discrete occurrence of transactional opportunities. The hypotheses are constructed to link predictive outcomes to managerial decisions related to targeting, budgeting, and the prioritization of interventions at the customer and segment levels. In a reproducible manner, the study proposes an operationalization of CLV, feature construction, temporal separation into calibration and validation periods, evaluation using error-based and classification performance metrics, analysis of explainability and calibration, and translation of analytical results into managerial decision rules.

The limitations of the study are conceptually related to the assumptions underlying probabilistic models, the latent nature of customer churn in non-contractual settings, and the scope of the applied demonstrations, which are structured as prototypes and scenario-based analyses. Specifically, the limitations arise from: (1) the dependence of comparative results on the assumptions of classical probabilistic models, including stationarity and parametric specifications, which may be violated under dynamic customer behavior, seasonality, or trends; (2) the characteristics of non-contractual contexts, in which churn is not directly observable and must be inferred indirectly through inactivity; and (3) the empirical fact that the analysis is based on simulated CRM-type datasets, albeit structured to closely resemble real-world data.

The empirical validation is organized around two concluding case studies, presented as appendices: a non-contractual omnichannel retail context and a contractual SaaS business model, through which prototyping and reproducible implementation in R are demonstrated.

II. Structure of the Dissertation

The exposition follows a hierarchical structure organized by chapters, sections, and subsections, in accordance with the internal logic of the research process. The dissertation comprises: Chapter 0 (Introduction), Chapter 1 (Conceptual and Economic Foundations), Chapter 2 (Typology of Models), Chapter 3 (Contractual and Non-Contractual Settings), Chapter 4 (Methodological Framework), Chapter 5 (Prototyping and Comparative Analysis), a Conclusion (entitled *Synopsis and Discussion*), as well as a Reflexive Epilogue, which presents selected authorial perspectives on the post-analytical research horizon and the transformation of economic thinking, followed by Appendix A and Appendix B. The main emphases of the individual parts are as follows:

- In **Chapter 0**, the motivation and the novel argumentative position of the study are developed, positioning CLV as a value-based framework reshaped by digitalization and machine learning, and emphasizing the need for direct comparisons between classical probabilistic and machine learning approaches under unified evaluation criteria.
- In **Chapter 1**, the economic logic of CLV and its strategic applications are elaborated, including segmentation, budgeting, and performance management, together with a system of related metrics and ethical considerations.
- In **Chapter 2**, a typology of CLV models is proposed, encompassing deterministic models, heuristic approaches, including RFM and rule-based methods, probabilistic Buy-Till-You-Die (BTYD) models, machine learning models, deep neural network approaches, and hybrid and ensemble constructions, accompanied by an analysis of assumptions, data requirements, advantages, and limitations.

- In **Chapter 3**, contractual and non-contractual customer relationships are distinguished, and the manner in which contextual conditions determine the observability of churn, the validity of modeling assumptions, and the selection of an appropriate predictive framework is formally specified.
- In **Chapter 4**, the methodological protocol is formulated, including the operationalization of CLV, data preparation, time-oriented validation, accuracy and uncertainty metrics, calibration, interpretability, and feature engineering.
- In **Chapter 5**, the applied core of the dissertation is implemented, encompassing probabilistic modeling, including BG/NBD and a monetary component, machine learning models, including Random Forest and XGBoost, explainability techniques, including feature importance, partial dependence, and SHAP, a hybrid framework based on stacking and calibration, and the derivation of managerial decision rules.
- **Appendix A** and **Appendix B** provide detailed protocols for the two case studies, ensuring analytical reproducibility and editorial symmetry between the non-contractual omnichannel retail context and the contractual SaaS context.

The full indexed table of contents is as follows:

III. Summary Of The Content And Results Of The Dissertation

CHAPTER 0. (Introductory Section)

In the introductory section (Chapter 0), the dissertation develops the argument for the strategic significance of Customer Lifetime Value, situates the concept within the context of digitalization, and formulates the methodological problem of integrating classical modeling approaches with machine learning. This chapter defines the object, subject, and scope of the research, as well as the objective framework and the hypotheses that establish the logic of the empirical testing.

CHAPTER 1. Essence and Economic Logic of the Customer Lifetime Value Concept

Chapter 1 fulfills a fundamental conceptual function within the dissertation. It firmly establishes the meaning of Customer Lifetime Value as an economic construct that simultaneously belongs to marketing, finance, and customer relationship management. The premise is unambiguous. Prior to the introduction of models, algorithms, and prototypes, it is necessary to clarify precisely what is meant by "value," how the "lifetime" horizon is defined, and what exactly is being forecast. This clarification is not pursued declaratively, but through a systematic examination of definitions, historical development, strategic applications, related metrics, model types, and ethical risks.

1.1. Definition of the Concept of CLV

This section argues that the literature contains multiple designations and "semantic variations" of the concept, including different abbreviations and usages that sometimes lead to pleonasms. Against this background, the chapter proposes a working core definition. Customer Lifetime Value is conceptualized as a quantitative indicator of the total monetary value that a customer generates for a firm over the duration of the relationship, most commonly formulated as the net present value of future cash flows.

A key scholarly contribution of this section is that the dissertation does not treat "value" as a given. On the contrary, the analysis emphasizes differences among definitions depending on whether value is measured as revenue, contribution margin, profit, or net present value, including differences in the underlying financial assumptions. The text develops an analytical categorization of definitions, distinguishing between approaches in which CLV is interpreted as a revenue-based measure and those that advocate a profit-based interpretation as a more appropriate foundation for managerial decision making. A further substantive contribution is the conceptual distinction between Customer Lifetime Value and Customer Equity, demonstrating that CLV is a micro-level construct, whereas customer equity represents the aggregated value of the customer portfolio.

To support this review, a table of CLV definitions is included, serving not only a descriptive but also a methodological function. It provides a reference point for subsequent discussions of modeling assumptions and evaluation metrics.

1.2. Evolution of the CLV Concept in the Context of Marketing Science

Section 1.2 traces the historical "maturation" of Customer Lifetime Value from early quantitative attempts toward a consolidated research program that integrates marketing strategy, financial valuation, and analytical modeling. The text highlights the role of seminal authors and publications that frame marketing as an investment and advocate the measurement of value through customer assets, including the work of Gupta and Lehmann, as well as discussions in the *Journal of Marketing* from the early 2000s that institutionalized the idea of "marketing return" through CLV and customer equity.

This section explicitly links the evolution of the concept to the development of methodological approaches. It emphasizes that by the mid-2000s, deterministic, probabilistic, and econometric strands had already been distinguished, while after 2010 the development of CLV modeling accelerated as a result of big data, cloud-based CRM systems, and automation. The section concludes by summarizing that Customer Lifetime Value has become a central axis of both theory and practice in customer-centric management, with a growing emphasis on ethical and regulatory considerations.

This historical account is not merely informative. It provides the rationale for the subsequent chapters' focus on synthesizing classical probabilistic models with machine learning approaches. In support of this argument, a table outlining the stages in the evolution of the concept is included, systematizing historical periods, key emphases, and methodological transitions.

1.3. The Strategic Role of Customer Lifetime Value in Customer Relationship Management, Customer Segmentation, and Marketing Budgeting

In this section, the dissertation makes a decisive transition from conceptualization to managerial application. Customer Lifetime Value is examined not as a purely statistical quantity, but as an instrument for customer prioritization, value-based segmentation, optimization of marketing expenditures, and the framing of strategic goal setting. Subsections 1.3.1 through 1.3.3 develop three interrelated lines of analysis.

- First, CLV is presented as a foundation for customer segmentation and targeting. The underlying logic is that segmentation is detached from purely descriptive criteria and anchored instead in economic value and future potential, enabling a more effective allocation of efforts toward customer retention and development.
- Second, CLV is examined as a basis for marketing budgeting and resource allocation. The section emphasizes that marketing budget decisions should be governed by expected value creation rather than by short-term performance indicators alone. In this sense, CLV functions as an economic "compass" for acquisition and retention expenditures.
- Third, CLV is discussed as an indicator of marketing effectiveness and strategic goal setting. Here, the dissertation underscores managerial accountability, the measurability of outcomes, and the possibility of integrating CLV into key performance indicator frameworks.

The analysis is supported by a table of the strategic applications of CLV, which structures decision types, required inputs, and expected managerial outcomes.

1.4. Key Metrics Related to Customer Lifetime Value

This section serves as a methodological bridge by identifying the metrics that are logically and empirically related to CLV and clarifying their implications for subsequent operationalization. The discussion encompasses customer acquisition cost, customer retention costs and their return, customer margin, return on investment, customer equity, and customer referral value. The analysis is critical in establishing that CLV cannot be meaningfully interpreted without explicit reference to margins and costs, and that additional "channels" of value, such as referral value, may extend the CLV framework beyond the immediate transactional revenue stream.

1.5. Methodological Approaches to Estimating and Forecasting Customer Lifetime Value

This section introduces the first systematic methodological framework of the dissertation, distinguishing four classes of approaches: deterministic, probabilistic, predictive approaches based on machine learning, and hybrid approaches. It establishes the methodological "contract" of the entire study, namely that different classes of models are appropriate under different data conditions, assumptions, and managerial objectives. A key feature of this section is the inclusion of a comparative table that summarizes the strengths and weaknesses, input requirements, and applicability of each approach. This table serves as a methodological guide for model selection in the applied part of the dissertation.

1.6. Ethical and Managerial Considerations

Chapter 1 concludes with a clearly structured ethical perspective. The use of CLV is shown to entail risks related to the commodification of customers, challenges concerning privacy and trust, and tensions between efficiency and ethics. The section discusses requirements for transparency, interpretability, and managerial accountability, which later motivate the inclusion of interpretable methods and diagnostic procedures in machine learning and Bayesian modeling. The discussion is supported by a table outlining ethical risks and corresponding mitigation guidelines, thereby conferring normative completeness and managerial relevance to the dissertation.

- **Scientific Result**

Within Chapter 1, a conceptual and methodological result is achieved that codifies Customer Lifetime Value as a micro-level economic construct, subject to financial interpretation and managerial operationalization. On the basis of a systematic differentiation of definitions and terminological usages, the chapter clarifies what is meant by “value,” how the “lifetime” horizon is interpreted, and which financial and behavioral components are admissible within the CLV framework. The historical tracing of the concept’s evolution within marketing science, combined with its strategic positioning in customer relationship management and customer-centric models, substantiates the treatment of CLV as an integrative indicator that links segmentation, targeting, budgeting, and the management of marketing effectiveness.

A key result is also the systematization of the “related” metrics that constitute the managerial infrastructure surrounding CLV, including customer acquisition cost, customer retention costs and their return, customer margin, return on investment, customer equity, and customer referral value. This systematization provides clarity regarding the boundaries of CLV as a metric and the conditions under which it can be applied as a criterion for resource allocation. The chapter concludes with a result of a normative nature, namely the formulation of ethical and managerial considerations that impose requirements for transparency, interpretability, and the responsible use of value-based models.

- **Relationship Between the Scientific Result, the Working Hypotheses (H1–H6), and the Empirical Validation**

The conceptual and economic result achieved in Chapter 1 provides the formal foundation for the testability of hypotheses H5 and H6. By defining Customer Lifetime Value as a value stream interpretable within a framework of long-term return and resource allocation, the chapter establishes the criterion of “managerial usefulness” that H5 requires to be validated not only statistically but also through its impact on decision making. At the same time, by systematically positioning CLV as a bridge between theory and practice, the chapter legitimizes hypothesis H6 concerning the appropriateness of hybrid solutions as a managerially rational synthesis.

From the perspective of empirical validation, Chapter 1 prepares the synopsis by firmly establishing what is being measured and what constitutes an improvement in model quality. This enables the results in Chapter 5 to be interpreted against clearly defined objectives rather than against arbitrarily selected metrics. The normative layer addressing transparency and accountability at the end of the chapter also supports the empirical interpretation of H2, as it introduces the requirement that interpretability be treated as a substantive criterion in the comparison between probabilistic and machine learning models.

- **Contribution**

The contribution of Chapter 1 lies in the construction of an editorially and methodologically robust "conceptual core" of the dissertation, which ensures unambiguous use of Customer Lifetime Value throughout the entire study. This core is realized through systematized definitions, a historical trajectory of conceptual development, and explicit links to managerial decision making, allowing subsequent models and protocols to be interpreted not as self-contained analytical exercises but as instruments for economically rational marketing interventions. An additional contribution is the integration of the ethical perspective as a formal component of the CLV framework, which strengthens both the scientific validity and the practical legitimacy of the methodological choices developed in the later chapters.

CHAPTER 2. TYPOLOGY OF MODELS FOR ESTIMATING CUSTOMER LIFETIME VALUE

Chapter 2 fulfills a systematizing and critical-analytical function. It builds upon the conceptual foundation developed in Chapter 1 by formulating a precise typology of CLV models grounded in their theoretical roots, assumptions, data requirements, mathematical formulations, and practical applicability. At the outset, it is stated that the typology encompasses deterministic models, heuristic and rule-based approaches, probabilistic (stochastic) models, machine learning models, deep neural network models, and hybrid models. Although the text includes an explicit enumeration, the underlying analytical intent is to demonstrate that the typology is not nominal but functional, in that different classes of models address different analytical and managerial problems.

2.1. Deterministic Models

Section 2.1 begins with the earliest and most intuitively accessible class of models. The exposition is organized around standard deterministic CLV formulas for infinite and finite horizons, tracing their origin to discounting and geometric progression, and emphasizing that these formulations assume temporal constancy of key parameters. The limitations are explicitly highlighted, including the absence of individual-level modeling, the inability to capture heterogeneity and dynamics, and a strong reliance on coarse averaging. Nevertheless, the dissertation demonstrates contexts in which deterministic models are justified, such as early stages of analytical maturity, highly aggregated data environments, or managerial communication settings where the objective is to provide a rapid directional assessment. The section concludes with a table of typical applications, positioning deterministic models within established industry practices.

2.2. Heuristic Models

Section 2.2 analyzes the RFM tradition and extended heuristic rule-based methods. These approaches are presented as an intermediate class, more informative than deterministic models yet lacking the strict generative structure of probabilistic models. The dissertation shows that heuristic approaches are widely adopted due to their simplicity, ease of interpretation, and direct linkage to CRM segmentation and targeting practices. A table of extended heuristic CLV models is included, summarizing rule types and extensions, such as the incorporation of additional behavioral indicators or threshold-based activation logic.

2.3. Probabilistic (Stochastic) Models

Section 2.3 constitutes a central pillar of the dissertation's theoretical apparatus. It examines Buy-Till-You-Die models, including Pareto/NBD and BG/NBD, as well as the logic of estimating purchase frequency and the probability that a customer is "alive," P(alive). The exposition integrates historical context, theoretical foundations, and practical applicability, emphasizing that these models are particularly relevant in non-contractual settings where churn is not directly observable.

To support intuition, the chapter includes figures illustrating the parametric components of the models and the behavior of the P(alive) function as a function of recency and frequency. Importantly, the chapter does not leave the reader solely with formal equations, but provides comparative tabular summaries, including a table comparing Pareto/NBD and BG/NBD and a table offering guidance for model selection between them. These summaries elevate the selection criterion from the level of "which model is better" to "which model is more appropriate under a specific behavioral and data context."

2.4. Machine Learning–Based Models

Section 2.4 positions machine learning approaches as a response to the limitations of classical probabilistic models, particularly their strong structural assumptions. The text examines the historical and theoretical roots of these approaches, their evolution from regression-based methods to more complex algorithmic frameworks, the core principles of ML-based CLV modeling, and their advantages and limitations. A central emphasis is placed on data requirements and the risk of "black-box" behavior, as well as on the necessity of interpretability and calibration. The section includes a table mapping ML-based algorithms to industries and data types, which has clear applied value by illustrating correspondences between data regimes, customer behavior patterns, and suitable algorithms.

2.5. Deep Neural Network–Based Models

This section treats deep neural network approaches as a distinct class due to their differentiated data requirements and their capacity to represent complex relationships. The discussion covers conceptual foundations, data requirements, advantages and limitations, conditions for application in business contexts, emerging innovative directions, and relevant software tools and libraries.

A key contribution of this section is the explicit articulation of the limits of practical usability. While DNN approaches are powerful, they require a mature data environment, appropriate infrastructure, and robust risk control procedures, including transparency and governance mechanisms.

2.6. Hybrid and Ensemble Models

Section 2.6 argues that in practice, the "optimal" solution is often hybrid. Hybridization is conceptualized as the purposeful combination of different classes of models, such as using probabilistic outputs as features in machine learning models, ensemble constructions, or combinations of probabilistic and causally oriented approaches to support managerial decision making.

The section includes subsections on forms of hybrid modeling, advantages and limitations, conditions for application, and software tools. It functions as a theoretical legitimization of the applied solutions developed in later chapters, where multiple models are combined and evaluated across multiple criteria.

2.7. Considerations in Selecting an Approach to CLV Modeling

The chapter concludes by systematizing model selection criteria through two comprehensive summary tables. One provides a comparative overview of typologies derived from the literature and practical assessment, while the other compares major CLV model types in terms of input data, assumptions, strengths, limitations, and software tools. This section consolidates the core methodological argument of the dissertation, namely that model choice is a function of context, data availability, forecast horizon, the need for interpretability, and organizational readiness.

- **Scientific Result**

Chapter 2 achieves a systematizing scientific result by developing a typology of CLV models encompassing deterministic models, heuristic approaches, including RFM and rule-based methods, probabilistic (stochastic) models, machine learning models, deep neural network models, and hybrid and ensemble constructions. The result extends beyond classification by providing an analytical differentiation based on theoretical foundations, assumptions, data requirements, strengths and weaknesses, and typical application scenarios.

A particularly significant result is the positioning of Buy-Till-You-Die probabilistic models as the reference class in non-contractual settings, while machine learning approaches are argued to be a flexible alternative when rich feature sets are available and when capturing nonlinearities, interactions, and complex behavioral patterns is required. The introduction of deep neural networks as a distinct category is justified by their higher data and infrastructure requirements, as well as by specific risks related to transparency and control. Hybrid and ensemble solutions are interpreted as a principled direction for synthesizing interpretability and predictive power.

- **Relationship Between the Scientific Result, the Working Hypotheses (H1–H6), and the Empirical Validation**

The scientific result of Chapter 2 establishes the direct logical framework for hypotheses H1, H2, H4, and H6. The differentiation between probabilistic and machine learning approaches in terms of assumptions, data requirements, and interpretability provides a theoretical justification for the expectation that machine learning models may outperform under conditions of rich data and high heterogeneity (H1), while probabilistic models exhibit structural advantages under limited data and strong assumptions (H4). In parallel, the typological positioning of probabilistic models as interpretable generative constructions supports hypothesis H2 regarding their superiority in calibration and managerial accountability.

From the perspective of empirical design, Chapter 2 supports the synopsis by defining which comparisons are scientifically admissible and what constitutes “comparable units” in the applied part of the study. This has direct implications for Chapter 5, where probabilistic, machine learning, and hybrid solutions are tested and interpreted as representatives of distinct model classes rather than as incomparable techniques. In this way, the chapter transforms H6 from a general proposition into a methodologically justified strategy for combining model classes that can be empirically evaluated.

- **Contribution**

The contribution of Chapter 2 lies in the creation of a unified methodological “map” of the CLV modeling landscape, which transforms a diverse set of approaches into a comparable system with clearly articulated selection criteria. This contribution is particularly important for the dissertation, as it legitimizes the subsequent head-to-head comparison between probabilistic and machine learning models and substantiates the necessity of hybrid constructions. The chapter also contributes by framing the software perspective, that is, by linking model types to practically implementable tools, which constitutes a prerequisite for the reproducibility of the applied developments in the subsequent sections.

CHAPTER 3. CUSTOMER LIFETIME VALUE IN CONTRACTUAL AND NON-CONTRACTUAL SETTINGS

Chapter 3 constitutes the contextual and analytical “hinge” of the dissertation. It introduces a key distinction that subsequently determines the methodological choices and applied protocols, namely contractual versus non-contractual customer relationships, as well as continuous versus discrete transactional opportunities. The central thesis is that these two contextual axes fundamentally alter the observability of churn, the meaning of recency and frequency, the structure of target variables, and the validity of modeling assumptions.

3.1. Typology of Customer Relationships

This section defines the two principal relationship types. In contractual relationships, customer status as “active” or “churned” is observable through an explicit contractual event. In non-contractual relationships, churn is latent and must be inferred indirectly through inactivity. The dissertation emphasizes that in the non-contractual scenario, the organization cannot determine with certainty the number of active customers, which explains why probabilistic Buy-Till-You-Die models are particularly relevant.

In parallel, a classification is introduced based on whether purchases can occur at any point in time or occur periodically. This distinction is methodologically decisive, as some models assume a continuous-time process, whereas others are discrete-time analogues.

3.2. Observability of Churn and Its Impact on CLV Modeling

This section formulates the direct relationship between churn observability and the construction of CLV models. In contractual settings, modeling can rely on explicit events and employ survival analysis, renewal models, and classification approaches. In non-contractual settings, probabilistic modeling of customer “liveness” is required.

3.3. Approaches to CLV Modeling in Non-Contractual Settings

This section examines probabilistic models for continuous-time purchasing processes, including Pareto/NBD and BG/NBD, models for discrete and periodic purchases as discrete analogues, monetary value estimation through a monetary component, and machine learning approaches.

A central emphasis is placed on the combined BG/NBD, or Pareto/NBD, framework for transaction frequency together with a Gamma-Gamma model for average transaction value, thereby yielding the expected monetary value of future purchases. This framework is presented as widely adopted in electronic commerce and is clearly positioned as a reference standard, which is subsequently employed in Chapter 5 during model prototyping.

3.4. Approaches to CLV Modeling in Contractual Settings

In this section, the dissertation examines survival analysis and churn modeling, models for predicting churn at renewal periods, and revenue forecasting in contractual contexts. This demonstrates how CLV in contractual settings can be more directly linked to churn risk, which is observable and subject to prediction. The section logically prepares Appendix B, where the contractual context is demonstrated through a SaaS scenario.

3.5. Typical Use Cases of CLV Models in Contractual and Non-Contractual Customer Relationships

Chapter 3 concludes with an application-oriented systematization of the types of decisions for which CLV models are used in the two relationship regimes. In support of this discussion, a comparative table of CLV models by context is included, along with two figures. One compares models in terms of complexity and interpretability, and the other presents a process diagram for model selection based on relationship type. This is a critical contribution, as it provides a methodological "decision tree" that eliminates arbitrariness in model selection.

- **Scientific Result**

Chapter 3 achieves a contextual and methodological scientific result by formalizing the differences between contractual and non-contractual customer relationships and demonstrating how these differences determine churn observability, the construction of target variables, and the validity of modeling assumptions. In contractual regimes, churn is observable and can be modeled as an event, whereas in non-contractual regimes, churn is latent and must be inferred probabilistically, necessitating specific models for estimating customer "liveness" and forecasting future transactions.

The result also includes a conceptual clarification of purchasing processes depending on whether transactional opportunities arise continuously or discretely, which is critical for model selection and temporal validation. In this way, the chapter functions as a theoretical bridge that translates the typology developed in Chapter 2 into the applicable protocols and models of Chapters 4 and 5, including by establishing the logic for distinct empirical settings in the appendices.

- **Relationship Between the Scientific Result, the Working Hypotheses (H1–H6), and the Empirical Validation**

The scientific result of Chapter 3, namely the contextual distinction between contractual and non-contractual relationships and its consequences for churn observability, is directly linked to hypotheses H1, H3, H4, and H6. The chapter explains why, in contractual settings, churn risk can be modeled as an observable event and therefore why approaches that explicitly predict churn and integrate it into value estimation should improve CLV assessment (H3). At the same time, it demonstrates why churn is latent in non-contractual settings and why models must be evaluated based on their ability to capture irregularity and heterogeneity, thereby supporting the formulation of H1.

With respect to empirical validation and the synopsis, Chapter 3 provides methodological legitimization for the two applied regimes developed in the appendices and reflected in the discussion. In this way, the hypotheses are anchored to context. Hypotheses H1 and H4 are primarily relevant to the non-contractual setting, whereas H3 pertains to the contractual setting. This transforms the empirical analysis in Chapter 5 and the appendices into a coherent testing of hypotheses across different relationship regimes, enabling the synopsis to summarize results without conflating non-comparable contexts.

- **Contribution**

The contribution of Chapter 3 lies in transforming the "relationship context" into an explicit methodological determinant of CLV modeling rather than a background characteristic. This provides a scientific foundation for methodologically sound comparisons among models, as comparability is defined with respect to relationship type and churn observability. On this basis, the chapter justifies why the same class of models may be optimal in one regime and inadequate in another, and why validation protocols and evaluation metrics must be selected contextually.

CHAPTER 4. METHODOLOGICAL FRAMEWORK FOR MODELING CUSTOMER LIFETIME VALUE

Chapter 4 constitutes the "methodological constitution" of the dissertation. It formalizes how a robust CLV analysis is constructed, specifying what exactly constitutes a model component, which assumptions are made, how data are prepared, which metrics are used for model evaluation, and where the "blind spots" of current scientific and applied practice reside.

At the outset, the chapter delineates the core elements of the framework, including a clearly articulated conceptual objective, specification of model components, formulation of assumptions, rigorous data preprocessing, selection of evaluation and benchmarking metrics, and identification of methodological gaps and innovations. This set of requirements establishes the standard according to which the prototypes in Chapter 5 are developed.

4.1. Components of CLV Models

In this section, CLV models are presented as analytical constructs that forecast the expected value of future cash flow streams. Core components are introduced, including the time horizon, purchase frequency and relationship duration, average transaction value, and margin. These components are subsequently organized at a higher level into operational, potential, relational, and contextual components.

The operational component is associated with measurable transactional quantities and cash flows. The potential component encompasses additional value derived from cross-selling, up-selling, and referral value, thereby extending CLV beyond immediate future purchases. The relational component incorporates parameters such as acquisition, retention, relationship duration, and accumulated customer experience. The contextual component accounts for environmental factors, channel structure, and product and business characteristics.

These concepts are supported by tabular descriptions of the components, visualizations of the scope of the research domains, and tabular representations of relationships and potential within the proposed models. Crucially, components are not treated solely at a conceptual level but are explicitly linked to measurement and modeling.

4.2. Core Assumptions in Customer Lifetime Value Modeling

This section crystallizes assumptions that are often implicit in CLV modeling, including independence of events, stationarity and temporal invariance, customer heterogeneity, and data granularity and sufficiency. These assumptions are directly relevant to later comparisons between probabilistic and machine learning models, as ML approaches may relax some of these assumptions while introducing other risks.

4.3. Challenges in Data Preparation and Preprocessing

Here, the dissertation systematizes practical obstacles, including sparsity of transactional data, truncation and censoring effects, information leakage, cohort and period effects, and the cold-start problem. This section establishes a concrete methodological discipline, emphasizing that validation and forecasting must be temporally correct and that features must be constructed so as not to "leak" information from the future.

4.4. Metrics for Evaluating and Comparing CLV Models

This section formalizes model evaluation across five dimensions: forecast accuracy metrics, including MAE, RMSE, MAPE, and R^2; ranking and classification performance metrics; return- and profit-based metrics; calibration of predictive estimates; and interpretability and transparency. This framework establishes the principle applied in Chapter 5 when comparing probabilistic, machine learning, and Bayesian approaches, namely that a model should not be evaluated solely on average error, but also on whether it correctly ranks customers and is managerially valid.

4.5. Methodological Challenges and Underexplored Areas

This section develops themes identified by the dissertation as critical for scientific advancement, including seasonality and temporal patterns, heterogeneity in discount rates and financial assumptions, time-varying covariates and behavioral dynamics, effects of customer satisfaction and experience, and real-time automation. It strengthens the scientific contribution of the dissertation by demonstrating not only what has been accomplished, but also the limitations of standard approaches.

4.6. Emerging and Prospective Research Directions

The final section of Chapter 4 positions research directions such as Bayesian calibration and hierarchical models, temporal embedding and deep learning, advanced survival analysis, causal uplift modeling for CLV optimization, and prospective unifying theoretical frameworks. This directly prepares the inclusion of a Bayesian prototype in Chapter 5 and supports the broader vision of the dissertation.

- **Scientific Result**

Chapter 4 delivers the central methodological result of the dissertation: a formalized framework for modeling, validating, and comparing CLV models, organized around CLV components, key assumptions, data preparation, evaluation metrics, and the identification of methodological challenges and emerging directions. The result lies in establishing a clear distinction between the operationalization of CLV and the statistical estimation procedures, while simultaneously enforcing temporal rigor through calibration and validation periods and the avoidance of information leakage.

A further significant result concerns model evaluation. A multi-criteria evaluation logic is established that does not reduce assessment to a single forecast error, but instead incorporates accuracy, calibration, uncertainty assessment, ranking effectiveness, and managerial interpretability. This enables Chapter 5 to conduct a comparison among probabilistic, machine learning, and Bayesian approaches within a unified methodological language.

- **Relationship Between the Scientific Result, the Working Hypotheses (H1–H6), and the Empirical Validation**

The scientific result of Chapter 4, namely the modeling protocol, time-oriented validation, and multi-criteria evaluation, constitutes the methodological prerequisite for empirically sound testing of all hypotheses, and most directly of H2 and H5. Hypothesis H2 requires distinguishing “accuracy” from “calibration” and interpretability, which is formalized through the separation of metrics and diagnostic procedures for calibration and transparency. Hypothesis H5 requires model evaluation through managerial impact, which is supported by the introduction of criteria related to ranking, targeting, and applied usefulness, rather than forecast error alone.

From the perspective of empirical validation, Chapter 4 enables the testing of H1 and H4 without methodological distortions such as information leakage or improper temporal partitioning. As a result, when probabilistic and machine learning models are compared in Chapter 5, the comparison occurs within a common validation framework, allowing the synopsis to draw conclusions regarding conditions of model superiority and limitations. In this way, Chapter 4 functions as a guarantee that empirical results are attributable to models and data rather than to procedural artifacts.

▸ Contribution

The contribution of Chapter 4 lies in the protocolization of a general framework that transforms the dissertation from a survey of models into a methodologically reproducible research program. The chapter consolidates the prerequisites for scientific testability of the hypotheses by defining how features are generated, how temporal validation is conducted, and how results are interpreted under uncertainty. An additional contribution is the identification of underexplored areas and emerging directions, positioning the dissertation as a contribution not only to applied methodology but also to the future research agenda.

CHAPTER 5. PROTOTYPING APPLIED MODELS FOR FORECASTING CUSTOMER LIFETIME VALUE

Chapter 5 constitutes the applied and empirical core of the dissertation. It operationalizes the methodological framework established in Chapter 4 through a series of prototypes that demonstrate: (1) data preparation and feature engineering, (2) probabilistic modeling with BG/NBD and a monetary component, (3) a machine learning–based CLV approach using a real transactional dataset, (4) a Bayesian CLV model as a generative alternative, and (5) a structured comparison of approaches using metrics, visual diagnostics, and validity checks.

5.1. Data Preparation and Feature Engineering

Section 5.1 establishes the foundation for applied modeling through systematic feature engineering. Four feature groups are examined.

- First, RFM metrics, namely recency, frequency, and monetary value, are treated as the classical core of transactional information.
- Second, tenure and relationship duration are introduced as an additional axis capturing customer “age” and stability.
- Third, cohort-based features are incorporated to account for acquisition-period effects and structural differences across customer groups.
- Fourth, sequence-based features are introduced to capture behavioral sequences, thereby extending the static RFM representation.

This section is critical because it ensures the methodological integrity of the subsequent prototypes while simultaneously demonstrating how classical and contemporary features can be combined.

5.2. Prototype of a Probabilistic BG/NBD Model

This section implements probabilistic modeling as the reference baseline, including its assumptions and structure, the construction of a customer-level model, the integration of customer heterogeneity, calibration, evaluation, and validation, comparison with Pareto/NBD and other model variants, the introduction of a monetary component via Gamma-Gamma and discounting, and forecasting.

Importantly, the dissertation employs the CDNOW transactional dataset, described as containing 78 weeks of purchase history for 2,357 customers, and demonstrates a reproducible implementation in R. Figures are included to visualize temporal purchase patterns, post-last-purchase scenarios, parametric densities of model distributions, and diagnostic comparisons between observed and predicted frequencies and transactions. Additional figures present the distribution of P(alive), a heat map of activity probability as a function of recency and frequency, the distribution of expected transaction counts, and the distribution of forecasted one-year CLV.

Tabular summaries are also provided, for example regarding approaches to forecasting monetary value, descriptive statistics of predicted CLV across segments, and a Pareto curve of CLV. These results facilitate a transition to managerial interpretation, emphasizing the concentration of value among a small fraction of customers and the implications for prioritization.

5.3. Prototype of a Machine Learning–Based CLV Model

Section 5.3 is organized into five logical steps: advantages and limitations, algorithms, procedure, forecasting, and comparative analysis.

- First, the dissertation argues why machine learning may be more appropriate under nonlinearities and interactions, while also emphasizing that it requires more data and stricter discipline to prevent overfitting and information leakage.
- Second, an overview of core algorithms for CLV is provided, including linear regression as a baseline and interpretable reference point, decision trees, ensemble methods, including Random Forest and gradient boosting using XGBoost, and neural networks. The discussion is not merely encyclopedic but emphasizes what each model gains and losses in terms of interpretability and predictive accuracy.
- Third, the procedure is demonstrated using the CDNOW dataset, including temporal partitioning into calibration and validation periods and the construction of features compatible with machine learning models. In the dissertation, this is implemented through a reproducible R script and a clearly described training and testing scheme.
- Fourth, forecasting using machine learning models is presented, distinguishing task types, including regression-based predictions and classification formulations, for example predicting active versus inactive status, which is directly linked to the problem of latent churn in non-contractual settings.
- Fifth, the comparative analysis includes predictive performance metrics, ROC curves for classification models, metrics for regression models, and graphical comparison of forecasting errors. Visual representations of decision trees are also included as part of the interpretability layer of the analysis.

5.4. Prototype of a Bayesian CLV Model

This section represents a high-level scholarly layer of the dissertation because it shifts from point estimation to a probabilistic generative perspective in which uncertainty is a structural component of the model and can be used for managerial decision making.

The section begins with problem formulation and motivation and then describes the Bayesian approach to CLV estimation and the generative structure of the core model. Strategies and algorithms for estimation follow, along with a concrete implementation of Bayesian estimation and forecasting, including a calibration and validation scheme.

The dissertation demonstrates a Bayesian implementation of BG/NBD and Gamma-Gamma components using brms, as well as the translation from posterior distributions to predictive estimates for P(alive), future transactions, and CLV. A significant contribution is the explicit conceptual linkage to decision making under uncertainty, focusing not only on “what is the expected value,” but also on “what is the probabilistic structure of risk and value,” which is relevant for threshold logic, budgeting, and risk management.

5.5. Comparison of the Bayesian Model with Reference Probabilistic and ML-Based Models

The final part of the chapter provides a methodologically disciplined comparison. It is not reduced to a single metric but combines multiple analytical elements.

- First, an analysis of the consistency of forecast results is conducted, including tabular comparisons of accuracy metrics.
- Second, visual inspection is performed, including correspondence plots between observed and predicted values, gain curves and cumulative accuracy, and distributions of forecasting error across models.
- Third, validity and reliability checks are implemented, including posterior predictive checks for the Bayesian layer and cross-validation for robustness.

In this way, Chapter 5 concludes with a clearly demonstrated principle, namely that model choice results from balancing accuracy, calibration, interpretability, and robustness rather than from reliance on a single performance statistic.

- **Scientific Result**

Chapter 5 delivers the empirical and analytical result of the dissertation through the prototyping of applied models for CLV forecasting, including probabilistic modeling with BG/NBD, extension with a monetary component (Gamma-Gamma) and discounting, prototyping of ML-based CLV models, and prototyping of a Bayesian CLV model, followed by a comparative analysis across approaches. The CDNOW transactional dataset serves as the empirical basis, on which the standard steps of BTYD modeling are demonstrated and a consistent protocol for training, forecasting, and validation is implemented.

The probabilistic prototype yields results including quantitative forecasts of the probability that a customer is active, the expected number of future transactions, and predicted CLV over a fixed horizon, supported by diagnostic checks and visualizations of consistency between observed and predicted frequencies and transactions. The ML prototype extends the framework to flexible algorithms for CLV forecasting, demonstrating a procedure for modeling and comparing results, including evaluation using appropriate metrics and analysis of predictive behavior. The Bayesian prototype adds a generative perspective and enables formal incorporation of uncertainty into forecasts and interpretation. The final comparative section consolidates the result through the comparison of probabilistic, ML, and Bayesian approaches under unified criteria.

- **Relationship Between the Scientific Result, the Working Hypotheses (H1–H6), and the Empirical Validation**

The scientific result of Chapter 5, namely the prototyping and comparative testing of probabilistic, ML, and Bayesian variants and their use for CLV forecasting and interpretation, constitutes the direct empirical test of hypotheses H1, H2, H4, and H6, and, at the managerial level, of H5. The comparison between the probabilistic prototype and the ML prototypes provides an empirical basis to evaluate whether ML yields improvement under heterogeneity and rich feature sets (H1) and whether probabilistic models retain advantages in calibration and interpretability (H2). The inclusion of scenarios with varying informational richness and dependence on structural assumptions supports the empirical evaluation of H4.

From the perspective of the synopsis, Chapter 5 enables the construction of conclusions that are simultaneously scientific and managerial. This is achieved by linking predictive quality metrics to actionable decision rules and by demonstrating a hybrid logic in which probabilistic forecasts and ML flexibility can be combined within an integrated framework (H6). Managerial verification of results, operationalized through the degree to which models support targeting and prioritization, provides an empirical basis for interpreting H5 in the concluding discussion as a criterion of practical validity for the scientific solutions proposed.

- **Contribution**

The contribution of Chapter 5 lies in demonstrating a complete, methodologically rigorous trajectory from data preparation and feature engineering through modeling and forecasting to diagnostics, comparison, and the managerial meaning of results, within a framework that ensures reproducibility in an R environment. The scholarly value of this contribution consists of three aspects. First, probabilistic BTYD models are implemented as a reference baseline against which a reasoned comparative argument is constructed. Second, machine learning and Bayesian approaches are included not as parallel exercises but as comparable alternatives evaluated under a common protocol and interpreted in light of their assumptions and limitations. Third, the comparative analysis is organized to support substantiated model selection as a function of data, context, and managerial requirements for transparency and reliability.

Synopsis and Discussion

The section “Synopsis and Discussion” synthesizes the comparative results of the empirical analysis conducted on the CDNOW dataset, organizing them around the three classical components of Customer Lifetime Value, purchase frequency, churn probability, and average monetary value per transaction. Within this integrative framework, three approaches are compared: (1) the frequentist BG/NBD plus Gamma–Gamma model estimated via maximum likelihood, (2) the Bayesian BG/NBD plus Gamma–Gamma model estimated via MCMC using Stan, and (3) a machine learning model based on Random Forest, trained on an identical set of behavioral and temporal features.

The comparison is structured along explicitly defined “key dimensions,” namely predictive accuracy, interpretability and informational content, scalability and operational complexity, and business context and use cases. In this way, the discussion establishes an analytical criterion for model selection that does not reduce model quality to a single metric, but instead evaluates it as a balance among accuracy, behavioral interpretability, resource feasibility, and contextual fit.

The methodological emphasis of the discussion is placed on the comparability of modeling setups and on the role of Bayesian specification as an added value in terms of uncertainty quantification and diagnostics, rather than as a guaranteed improvement in pointwise predictive error. In this sense, the section explicitly distinguishes between “point accuracy” and “uncertainty management,” emphasizing that the Bayesian BG/NBD replicates the point accuracy of the frequentist model while adding uncertainty information that is particularly valuable in settings with sparse data or small cohorts.

Methodological Syntheses Relevant to Validation

A central methodological conclusion of the discussion concerns the disciplined comparability of modeling assumptions and the analytical role of Bayesian inference. Bayesian specification is positioned not as a universally superior predictive solution, but as a framework that enhances diagnostic insight and uncertainty awareness. This distinction reinforces the idea that predictive performance should be interpreted jointly with uncertainty structure, especially in applied decision-making contexts.

Empirical and Analytical Results

In a dedicated subsection titled "Synthesis of Results," it is reported that the Random Forest model achieves the highest predictive accuracy, with R^2 greater than 0.84 and RMSE approximately equal to 1.0 under five-fold cross-validation. In contrast, the frequentist and Bayesian BG/NBD models exhibit similar and stable error levels, with RMSE in the range of approximately 1.767 to 1.8 and MAE around 0.81, alongside high internal consistency between the two parametric variants. For the Bayesian specification, adequate chain convergence and satisfactory posterior predictive checks are documented.

At the customer level, the discussion identifies differences in calibration. For highly active customers, Random Forest tends to overestimate expected value, whereas BG/NBD remains conservative. For low-activity customers, probabilistic models yield more realistic forecasts and exhibit lower sensitivity to noise. These calibration patterns are emphasized as critical for managerial interpretation and risk-sensitive decision making.

Interpretation and Methodological Implications

In the subsection "Prospective Methodological Implications and a Hybrid Framework," the results are interpreted through the thesis that accuracy and interpretability are complementary rather than mutually exclusive dimensions. The discussion articulates the position that no universal "winner" exists. Probabilistic models provide a transparent and structurally consistent foundation, whereas machine learning models deliver superior predictive power when rich feature sets are available. On this basis, a hybrid perspective is advanced, combining the interpretability of BG/NBD with the predictive flexibility of machine learning and the uncertainty intervals provided by Bayesian inference.

This perspective is operationalized in the text as a promising hybrid framework consisting of the following sequence. First, baseline calibration and segmentation are conducted through fitting BG/NBD plus Gamma–Gamma models to extract interpretable parameters and establish initial segmentation. Second, predictive refinement is achieved via machine learning by training a model with rich features to predict CLV over predefined horizons of three, six, or twelve months. Third, integration is implemented through stacking or ensembling, whereby BG/NBD outputs or posterior summaries are used as inputs to machine learning models, or through dual-output modeling aligned with probabilistic logic. Fourth, uncertainty management is addressed through Bayesian intervals for key segments or cohorts to support budgeting and risk management, including the use of variational or maximum a posteriori approaches in large-scale data settings. Finally, operationalization is ensured through a regular re-estimation cycle, interpretability via SHAP or partial dependence methods, and alignment with business-facing key performance indicators, including CAC, ROMI, and margin-based CLV.

- **Scientific Result**

The scientific result of the "Synopsis and Discussion" section lies in the synthesized demonstration of comparative validity between frequentist and Bayesian BG/NBD models and a machine learning approach based on Random Forest under an identical feature set, with particular emphasis on the distinction between pointwise predictive accuracy, customer-level calibration, and the value of uncertainty as an analytical resource.

- **Relationship to Hypotheses H1–H6 and the Function of the Synopsis**

The section consolidates the empirical testability of the hypotheses concerning differences between machine learning and probabilistic approaches in a non-contractual context, explicitly documenting the superiority of machine learning in terms of predictive accuracy while simultaneously demonstrating that calibration by customer profile and uncertainty management remain critical for practical applicability. Within the synopsis, this is reflected as an argument for conditional model selection and for methodologically justified combination of approaches, rather than reduction to a single model or a single metric.

- **Contribution**

The contribution of this section consists in translating empirical differences among approaches into explicit selection criteria and into a methodologically grounded perspective on hybridization. This perspective integrates structural interpretability, predictive power, and uncertainty management within a logic suitable for subsequent operationalization.

EPILOGUE: CUSTOMER LIFETIME VALUE IN THE ERA OF ARTIFICIAL INTELLIGENCE AND ANALYTICAL CAPITALISM

The section “Reflexive Epilogue” serves a concluding interpretative function by situating Customer Lifetime Value within the broader context of the digital economy, artificial intelligence, and analytical capitalism. Here, CLV is articulated not as a technical metric, but as a conceptual axis of a new marketing paradigm in which data and probabilities replace classical market dimensions.

Key Concepts and Conceptual Emphases

In the subsection “The Paradigm Shift in the Understanding of Value,” value is conceptualized as a potential, forecasted, and probabilistically assessable outcome of interactions between the customer and the organization. In this context, CLV is defined as an epistemological model, a way of thinking about and managing economic reality through predictability, in which the customer is conceptualized as a dynamic vector of probabilistic expectations whose value evolves over time and depends on forecasted activity.

In the subsection “Artificial Intelligence as a New Mediator of Economic Knowledge,” the epilogue introduces machine prediction as a form of knowledge that is operational in nature, meaning that it does not merely describe reality, but models and constructs it. The emphasis is placed on the idea that CLV models, particularly in their Bayesian and machine learning–based forms, embody this transformation by not only estimating value, but generating possible future behavioral scenarios.

In the subsection “Analytical Capitalism and the Customer as a Statistical Entity,” the argument is advanced that a new form of economic power emerges, namely control over prediction, whereby CLV models transform data into predictions and predictions into strategic decisions. In this logic, the customer is represented as a distribution of probabilities, managed through forecasts and algorithmic mechanisms.

Ethical Reflection and the Limits of Prediction

The epilogue raises the question of the ethical boundaries of prediction, emphasizing the risk of reducing human interaction to model-based abstraction. The responsibility of the researcher and the analyst is framed as extending beyond the pursuit of predictive accuracy. It includes reflection on the consequences of model-based knowledge, how it is used, whom it benefits, and whom it excludes. In this sense, CLV modeling emerges simultaneously as an analytical instrument and an ethical trial, a boundary between knowledge and power, and between forecasting and control.

Concluding Perspective

In the subsection “Toward a Post-Analytical Horizon: Knowledge, Time, and Value,” the modeling of Customer Lifetime Value is interpreted as a metaphor for the transformation of economic thinking, from retrospective analysis to prospective modeling. CLV is defined as a tool through which economic rationality is projected into time, and value is conceived not as a fact, but as a possibility. Under this perspective, every forecast of customer value constitutes both a scientific act and a social construction.

- **Scientific Result**

The scientific result of the epilogue lies in the conceptual positioning of CLV as an analytical axis of the predictive economy and as an epistemological framework that links value, time, and uncertainty within a logic of algorithmic decision making and probabilistic governance.

- **Relationship to Hypotheses H1–H6 and the Function of the Synopsis**

The epilogue supports the synopsis by providing a conceptual and ethical framing of the criteria that the dissertation identifies as central to the applicability of models, not only accuracy, but also transparency, interpretability, resource feasibility, and the manageability of uncertainty. In this way, the concluding reflection strengthens the argument that model selection and integration are inseparable from responsibility and control over the consequences of predictive knowledge.

- **Contribution**

The contribution of this section consists in legitimizing methodological and managerial requirements for transparency, interpretability, and reflexive consideration of the consequences of model-based knowledge as indispensable elements of the use of CLV models in contemporary data-rich and algorithmic environments.

APPENDIX A: OMNICHANNEL FASHION RETAIL, NON-CONTRACTUAL SETTING

Appendix A presents a reproducible analytical protocol for modeling Customer Lifetime Value in a non-contractual retail environment characterized by omnichannel interactions, irregular purchasing behavior, and latent churn, meaning that churn is not observed as a clearly defined event. The methodology is conceptually aligned with Sections 3.3 and 5.2–5.3 and is grounded in the logic of the behavioral process of repeat purchasing and the distinction between “temporary inactivity” and “true relationship termination.” Within this context, the primary methodological challenge is the correct specification of the time horizon, the appropriate separation of training and forecasting periods, and the control of information leakage risk, in direct correspondence with Sections 4.3.3 and 4.3.

The protocol begins with the formalization of transactional granularity and the construction of stable customer behavioral indicators, including recency, frequency, and monetary value, as well as omnichannel features that capture interactions among channel, product, and behavior. This is followed by probabilistic modeling of repeat purchases using Buy-Till-You-Die–type models as a reference baseline and subsequent estimation of the monetary component in order to derive a forecast of future customer contribution over a specified horizon. On this basis, managerial interpretations are developed in alignment with Section 1.3, including customer ranking by value, segmentation by risk and value, and the formulation of actionable targeting and budgeting rules in an omnichannel context. A distinctive feature of the methodology is that it presents not only model-based estimates, but also the full logic of validation, calibration, and sensitivity analysis with respect to key assumptions such as stationarity and heterogeneity, which are central to Sections 4.2 and 4.5.

APPENDIX B: SAAS SUBSCRIPTION MODEL, CONTRACTUAL SETTING

Appendix B presents a traceable and reproducible protocol for modeling Customer Lifetime Value in a contractual SaaS business environment, in which customer retention and churn are directly observable and the data exhibit natural periodicity defined by renewal and billing cycles. The methodology is directly aligned with Sections 3.4 and 5 and treats CLV as a function of the probability of retention in future periods and the expected conditional value given retention. In this context, the central methodological task is to correctly define customer states over time, select an appropriate forecasting horizon, and construct a validation scheme that closely mimics a real-world deployment scenario, in accordance with Sections 4.3 and 4.4.

The protocol begins with a clear operationalization of customer churn in discrete time intervals, the definition of training and forecasting periods, and the construction of SaaS-specific features, including plan characteristics, product usage, relationship tenure, and behavioral dynamics. This is followed by churn risk models and revenue forecasting components, evaluated not only in terms of predictive error but also with respect to calibration at the aggregate level, since managerial decisions are typically made under budget constraints through the selection of top segments and intervention thresholds. The methodology culminates in the translation of forecasts into managerial tools, such as a risk–value matrix, rules for differentiated interventions, and simulation of the expected impact of targeting strategies. These elements are directly linked to Sections 1.3.2, 4.4.3, and 4.4.2. In this way, the case functions as a complete procedural framework for CLV implementation in a SaaS context, including reliability control, robustness assessment of results, and evaluation of the trade-offs among predictive accuracy, interpretability, and organizational feasibility.

IV. Statement of Scientific Contributions of the Dissertation

The scientific contributions claimed in the present dissertation are as follows.

1) **Conceptual and historical consolidation of Customer Lifetime Value as an economic and managerial construct.**

   The dissertation provides a comprehensive review and historical synthesis of the development of Customer Lifetime Value, linking the theoretical foundations of CLV to the evolution of methodological approaches and to practices of customer-centric management. Within this conceptual line, CLV is formalized as a value stream relevant to strategic decisions related to segmentation, targeting, and budgeting, with a clearly articulated managerial logic connecting marketing investments to long-term profitability.

2) **Development of a systematic typology of CLV models and criteria for methodologically grounded model selection.**

   The dissertation proposes a typology encompassing deterministic, heuristic, probabilistic (stochastic), machine learning–based, deep neural network–based, and hybrid and ensemble CLV models. For each class, assumptions, data requirements, applicability, advantages, and limitations are explicitly distinguished. This results in a scientifically significant contribution that transforms a diverse set of approaches into a comparable system suitable for reasoned model selection based on context, data availability, and managerial objectives.

3) **Contextual framework for CLV in contractual and non-contractual relationships and its operational implications for modeling.**

   The dissertation formalizes the distinction between contractual and non-contractual customer relationships, including its consequences for churn observability and model specification. This contribution is methodologically substantive, as it defines the conditions under which probabilistic models constitute a natural choice due to latent churn, as well as the conditions under which contractual regimes allow direct modeling of customer churn as an observable event and its integration into value estimation.

4) **Protocol-based methodological framework for CLV modeling and validation, oriented toward time, calibration, and interpretability.**

   The dissertation formulates a coherent framework that includes CLV operationalization, data preparation, feature engineering, and time-oriented validation, accompanied by multi-criteria model evaluation. Within this framework, predictive accuracy, calibration, and uncertainty management are explicitly distinguished, along with requirements for interpretability and business applicability. This provides a methodologically sound foundation for subsequent comparative testing and for the derivation of robust managerial conclusions.

5) **Empirical comparison of probabilistic and machine learning approaches on a real dataset under unified validation conditions.**

   In the applied core of the dissertation, a comparative evaluation is conducted between frequentist BG/NBD and Bayesian BG/NBD models augmented with a monetary component, and a Random Forest model as a machine learning approach, using a comparable evaluation protocol. The discussion synthesizes results in terms of predictive accuracy, error behavior, and calibration across customer profiles, empirically demonstrating the trade-off between predictive power and structural transparency and deriving model selection criteria that are valid within the dissertation's framework.

6) **Methodological validation of the Bayesian perspective as a means of uncertainty management and diagnostic assessment of CLV model adequacy.**

   The dissertation integrates a Bayesian implementation of probabilistic modeling, emphasizing the role of posterior distributions, uncertainty intervals, and posterior predictive checks in assessing the adequacy and stability of forecasts. This contribution is scientifically relevant, as it expands the focus from point predictions to probabilistic assessment, which is applicable to decision making under risk and in settings characterized by sparse data or small cohorts.

7) **Reproducible applied protocols in two contextual settings and reflexive framing of CLV in the era of artificial intelligence.**

   Appendix A and Appendix B present traceable and reproducible protocols for CLV modeling in a non-contractual omnichannel retail environment and in a contractual SaaS setting, respectively, including clearly documented data logic, operationalizations, and modeling procedures consistent with the methodological framework of the dissertation. The final "Reflexive Epilogue" complements the scientific contribution by conceptually positioning CLV as an analytical and managerial artifact in the context of artificial intelligence and analytical capitalism, thereby legitimizing the importance of transparency, accountability, and responsible use of predictive knowledge as integral components of the scientific and practical validity of CLV modeling.

V. List of Prior Publications Related to the Topic of the Dissertation

Кръстевич, Т. (2026). Пожизнената стойност на клиента. Библиотека „Стопански свят“, АИ Ценов, Свищов (монография, под печат).

Кръстевич, Т. (2026). Предиктивно и каузално моделиране на мултитъч атрибуция в дигитална среда. Алманах "Научни изследвания", 34, АИ Ценов, Свищов, (студия, под печат)

Krastevich, T. (2026). Beyond Heuristics: A Predictive Modeling Framework for Multi-Touch Attribution in Online Marketing. Conference proceedings ICBE 2025. Bucharest, Romania: Springer. (SCOPUS, accepted)

Krastevich, T., & Vassileva, B. (2026). Robustness and Reliability of AI-Generated Personas in Customer Analytics. Conference proceedings FAIEMA 2025. Stavanger, Norway: Springer. (SCOPUS, accepted)

Krastevich, T., Vassileva, B., & Aleksandrova, Y. (2026). Decoding purchase sequences: Identifying patterns in the customer journey. Conference proceedings ICBE 2025. Bucharest, Romania: Springer. (SCOPUS, accepted)

Krastevich, T. (2024). A methodology for data-driven multi-touch channel attribution in online settings. In ICEBM, 26-27 September, Varna, pp. 10.-21, DOI: https://doi.org/10.56065/ICEBM2024.10

Krastevich, T. (2023). Retargeting Customers Using Uplift Modeling. Economic Studies. *Ikonomicheski Izsledvania*, 32(2), 78-99. (студия, Scopus)

Кръстевич, Т. (2022). Разкриване на закономерности при пазаруване в среда на големи данни. Алманах "Научни изследвания", 30, АИ Ценов, Свищов, стр. 138-180.

Krastevich, T. (2013). Using predictive modeling to improve direct marketing performance. *Ikonomicheski Izsledvania*, 22(3), pp. 25–55 (Scopus)

Кръстевич, Т. (2023). Клъстеризиране на домакинствата въз основа на заявена склонност към устойчиво потребление. В „Устойчиво потребление в градска среда – регионални различия“. ВСУ „Черноризец Храбър“, Варна, стр. 167-210, ISBN 978-954-715-754-4

Кръстевич, Т., Смокова, М. (2015). Предиктивен анализ на потребителските реакции чрез инкрементално моделиране. Алманах "Научни изследвания", 22, АИ Ценов, Свищов, стр. 352-385.

Кръстевич, Т., Смокова, М. (2014). Прогнозиране на продажбите на бързооборотни потребителски стоки на базата на данни от продажбени трансакции с вероятностни модели. Алманах "Научни изследвания", 21, АИ Ценов, Свищов, стр. 89-116.