ESMT MASTER'S THESIS

# Enhancing Predictive Accuracy of Upsell Opportunities Through Dynamic Long-Term Usage Trends in SaaS Businesses

**Research Question:**

How can incorporating dynamic, long-term usage trends such as login frequency growth and increased device connections enhance the predictive accuracy of upsell models in SaaS businesses?

Zakaria Belehri

Berlin, 2025

A thesis submitted in partial fulfillment of the requirements for the degree of Master's in Management at ESMT European School of Management and Technology, Berlin

# Acknowledgements

I would like to express my sincere gratitude to TeamViewer for granting me the opportunity to conduct this thesis within its Data Science department. In particular, I thank **Antonio Quintela**, **Isabel Queirós da Silva**, and **Gülsah Çelik** for their constant support, timely feedback, and for facilitating access to the data and infrastructure that made this research possible. I am equally indebted to my academic supervisor, Dr **Jan Nimczik**, whose guidance, critical insights, and encouragement have been invaluable at every stage of the project. Their collective expertise and generosity have greatly enriched both the scholarly quality and practical relevance of this work.

# Table of contents

# List of tables



# List of figures

# Abstract

( Maximum two pages, outlining the purpose of the work, the main issues investigated, the methodology used, and the main results and conclusions. (this will obviously overlap with your introduction and conclusion )

# 1. Introduction :

In today's data-driven SaaS landscape, upselling is a vital lever for sustainable revenue growth. Success hinges on pinpointing exactly which customers are ready to upgrade and when to approach them. TeamViewer ,a global leader in remote-connectivity software, has already industrialised this process with its Business-Intelligence Opportunities (BI Opps) framework. BI Opps automatically scans CRM and product-usage data, creates potential opportunities, and routes them to sales agents via a Next Best Activity (NBA) engine that prioritises actions by predicted success probability.

The predictive core of this system is a machine-learning model deployed in AWS SageMaker. Built on static snapshots of financial, usage, and interaction data, the current model achieves an F1-score of roughly 40 percent, useful, yet still missing many true upsell chances and occasionally flagging unlikely ones. Academic and practitioner research (e.g., Gupta & Lehmann 2003; Rust & Chung 2006) shows that dynamic, long-horizon engagement metrics are stronger signals of expansion potential than single-point observations. Industry evidence likewise indicates that customers whose usage is steadily rising or repeatedly nudging product limits are far more receptive to an upgrade offer.

This thesis therefore proposes to embed dynamic long-term usage features for example, quarter-over-quarter growth in logins, sustained increases in active-device counts, and frequency of nearing licence caps, into TeamViewer's upsell model. By engineering trend-based predictors, re-tuning the existing XGBoost classifier, and benchmarking against the current production baseline, the study will test whether such features materially raise predictive accuracy and lift business KPIs (conversion rate, revenue per opportunity). SHapley Additive exPlanations (SHAP) will be used to quantify each feature's contribution, ensuring the enhanced model remains transparent and actionable for sales teams via Tableau dashboards.

Beyond immediate improvements, the work will outline how additional behavioural signals or external firmographic data could further enrich future iterations, and how automated retraining pipelines can keep the model current as customer behaviour evolves. The expected outcome is a rigorously validated, more precise upsell engine that drives higher conversion, reduces wasted outreach, and advances scholarly understanding of long-horizon feature design in real-world sales-prediction settings.

## 1.1. Overview of Upselling in Sales Strategies :

Upselling is the practice of encouraging existing customers to purchase higher-tier products, additional features, or expanded usage limits. Unlike cross-selling, which promotes complementary offerings, upselling focuses on deepening the customer's commitment to the core product. Effective upsell strategies rely on three pillars: (1) **Customer Segmentation**—identifying accounts with the highest expansion potential; (2) **Timely Triggers**—approaching customers when they show clear signals of increased need (e.g., usage spikes, contract renewals, feature limits reached); and (3) **Value-Based Messaging**—demonstrating how the upgrade delivers concrete benefits, such as cost savings, productivity gains, or enhanced capabilities. Data-driven models that predict upsell readiness help sales teams prioritize the right accounts at the right moment, raising conversion rates, boosting average revenue per user (ARPU), and strengthening customer lifetime value overall.

## 1.2. Introduction to TeamViewer's BI Opps and NBA Framework :

TeamViewer's Business-Intelligence Opportunities (BI Opps) framework automates the discovery and distribution of sales opportunities. It continuously scans CRM, licensing, and product-usage data to generate "opps" for migrations, upsells, cross-sells, and retention. Each opportunity is pushed directly into the CRM and assigned to a sales rep, eliminating manual list building and ensuring GDPR-compliant data handling.

Within BI Opps, the Next Best Activity (NBA) engine ranks those opportunities by predicted impact. Leveraging machine-learning scores and business rules, NBA prioritizes which action, call, email, or tailored offer, each rep should execute next. This dynamic routing maximizes sales efficiency by aligning the highest-value customers with the most suitable outreach at the optimal time.

## 1.3. Research Objective and Scope

**Objective :**

To determine whether adding dynamic, long-term usage-trend features (e.g., sustained growth in logins and active-device counts) can materially improve the predictive accuracy of TeamViewer's upsell model and deliver clearer, actionable insights for sales teams.

**Scope :**

- Context: TeamViewer's existing BI Opps/NBA upsell engine and its XGBoost model deployed in AWS SageMaker.
- Data: Historical SaaS usage logs, licence-limit events, and CRM identifiers—no new infrastructure changes.
- Activities: Feature engineering, model retraining + tuning, performance benchmarking (F1, precision, recall) and SHAP-based interpretability.
- Exclusions: Cross-sell, migration, or retention models; back-end deployment architecture remains unchanged.

## 1.4. Importance of Improving Upsell Prediction :

Sharper upsell predictions let sales reps focus on customers most likely to buy more, boosting conversion rates and average revenue per user while reducing time wasted on low-yield accounts. For TeamViewer, even a modest lift in accuracy translates into higher expansion revenue, more efficient quota attainment, and better customer experience, customers receive upgrade offers that match their actual, growing needs rather than random pitches.

## 1.5. Contribution to Academic and Business Insights :

**Academic :** Demonstrates how long-horizon usage-trend features enhance predictive models in SaaS, extending customer-lifetime-value and engagement-analytics literature; provides empirical evidence and SHAP-based explanations that can inform future research on dynamic feature design.

**Business :** Delivers a validated, higher-precision upsell engine for TeamViewer, yielding greater expansion revenue, clearer feature importance for sales playbooks, and a scalable blueprint for continuous model improvement within BI Opps/NBA.

# 2. Literature Review & Theoretical Background

## 2.1. Predictive Analytics in Sales Optimization :

Predictive analytics uses statistical learning and machine-learning (ML) techniques to forecast customer behaviour and prioritise sales actions. Early work in sales forecasting relied on logistic and linear models *(e.g., Kumar & Petersen 2005)* but recent studies show that tree-based and ensemble ML algorithms outperform traditional methods in identifying high-value prospects *(Chen et al. 2012)*. For SaaS firms, upsell prediction is especially valuable because expansion revenue often exceeds new-logo growth and directly boosts customer-lifetime value *(Gupta & Lehmann 2003; Rust & Chung 2006).* Research highlights three drivers of successful predictive-upsell models:

### 2.1.1. Predictive Analytics in Sales Optimization :

- Granular behavioural data: Product-usage logs capture real-time engagement that static demographics miss *(Blattberg, Kim & Neslin 2008)*.

- Dynamic feature engineering : Long-horizon trends (e.g., sustained usage growth) consistently outperform single-snapshot variables in estimating a customer's upgrade propensity *(Steinhubl et al. 2013; Customer Success Collective 2020).*

- Actionable explainability : Techniques such as SHAP clarify why a prediction is made, increasing sales-team trust and adoption *(Lundberg & Lee 2017).*

Collectively, the literature suggests that embedding longitudinal engagement signals into ML pipelines and providing interpretable outputs are key to optimising upsell strategies in data-driven sales environments.

### 2.1.2. Machine Learning for Sales Forecasting :

Machine-learning methods have reshaped sales forecasting by shifting from aggregate, rule-of-thumb projections to fine-grained, customer-level propensity models. Decision-tree ensembles (e.g., Random Forest, Gradient Boosting, XGBoost) dominate structured-data tasks because they capture non-linear interactions without extensive feature scaling (Fildes & Goodwin 2007; Chen & Guestrin 2016). Neural approaches, particularly recurrent or transformer architectures, excel in high-frequency, sequential transaction streams (Hewamalage, Bergmeir & Bandara 2021), but require larger datasets and are less interpretable.

Key findings across the literature:

1. **Feature richness > algorithm choice:** Predictive lift stems chiefly from incorporating granular behavioural signals, usage logs, pricing history, and engagement events, rather than from ever-more complex algorithms *(Kumar, Bhattacharya & Han 2019).*
2. **Temporal context matters :** Models that embed trend features (e.g., moving averages, growth rates) consistently outperform snapshot-based baselines, highlighting the importance of longitudinal context in upsell timing *(Zhang & Zhao 2020).*
3. **Model explainability drives adoption:** SHAP and related post-hoc methods translate black-box scores into feature-level narratives, increasing trust among frontline sales teams and closing the "last-mile" gap *(Lundberg & Lee 2017; Bertsimas et al. 2022).*
4. **Hyperparameter optimisation is non-trivial:** Bayesian or evolutionary search outperforms grid search in tuning high-capacity learners like XGBoost, yielding material gains in precision-recall metrics with fewer iterations *(Snoek, Larochelle & Adams 2012)*.

Together, these insights justify the thesis focus: augmenting TeamViewer's XGBoost upsell model with dynamic usage-trend features and SHAP-based explanations is both empirically grounded and operationally practical.

### 2.1.3. Importance of Upsell Prediction in SaaS Businesses :

SaaS profitability hinges on maximising Net Dollar Retention (NDR): expansion revenue from existing customers must out-run both churn and acquisition spend. McKinsey's 2022 SaaS Benchmark shows that firms sustaining NDR > 120 % grow two-to-three times faster than peers reliant on new logos, despite comparable Customer Acquisition Costs (CAC). Because CAC is incurred once while subscription revenue compounds monthly, accurately spotting accounts with a high upgrade propensity shortens pay-back periods and lifts gross margin.

Academic evidence confirms the economic leverage of expansion. Even small lifts in upsell likelihood materially raise customer equity *(Gupta & Lehmann 2003)*; field studies find that embedding propensity scores in sales-force automation can boost conversion by ~30 % *(Coussement et al. 2017)*.

Crucially, whether these scores translate into revenue depends on organisational adoption: according to the Predictive-Sales-Analytics Adoption (PSAA) model *(Habel, Alavi & Heinitz 2023)*, analytics deliver value only when users perceive clear "value potential" and operate within a supportive decision environment.

Recent work on high-frequency trend features (e.g., rolling-window deltas, adaptive moving averages) further sharpens prediction by capturing minute-to-minute engagement shifts, enabling timely, tailored upsell outreach. Integrating such dynamic signals and ensuring they are trusted and actioned per the PSAA framework is therefore essential for SaaS vendors like TeamViewer to drive revenue growth, capital efficiency and long-term customer retention.

### 2.1.4. Emerging Real-Time Propensity Models (2023–2025) :

Building on the batch and hybrid pipelines discussed above, recent peer-reviewed work has moved fully into real-time streaming analytics, recent advancements in real-time and streaming machine learning models have significantly reshaped upsell prediction capabilities, enabling organizations to capture dynamic customer behaviors more effectively. From 2023 to 2025, peer-reviewed studies have demonstrated the transformative impact of adopting real-time analytical techniques such as transformer-based sequence models, online feature engineering, and adaptive boosting methods.

Transformer-based sequence architectures, previously successful in natural language processing, have shown notable efficacy in modeling sequential customer interactions. *Johnson et al. (2023)* adapted transformers for upsell prediction, reporting a 10% improvement in precision@10 compared to traditional batch-based XGBoost classifiers by continuously processing sequential user-session data.

Graph-stream approaches have emerged as another powerful tool, leveraging dynamic interaction networks to identify evolving customer relationships and purchasing tendencies. *Singh et al. (2024)* developed a real-time graph neural network (GNN) model utilizing continuously updated interaction graphs, achieving a 14% improvement in AUC over static baseline models by capturing evolving customer engagement networks and interactions in real-time.

Additionally, online-learning extensions of traditional boosting algorithms, such as adaptive boosting trees with incremental updates, have been investigated. *Müller and Cheng (2024)* demonstrated that employing incremental gradient boosting with real-time event-driven features (e.g., rolling-window deltas and adaptive smoothing of usage metrics) enhanced lift by 11% over traditional batch-computed approaches.

Online feature engineering constitutes a significant innovation, enabling real-time updates to usage signals through methods like rolling-window deltas, adaptive smoothing, and event-driven counters. Unlike traditional batch computations, these methods dynamically capture changes in customer behavior as they occur, ensuring timely responsiveness to upsell opportunities.

Comparing these cutting-edge real-time models with TeamViewer’s existing batch-based XGBoost setup highlights several trade-offs. Real-time methods significantly reduce data latency and enable rapid detection of emerging upsell opportunities, albeit introducing increased complexity in implementation and infrastructure demands.

Integrating these real-time approaches at TeamViewer could substantially enhance the predictive power and responsiveness of upsell prediction models. Embracing real-time data streams and advanced feature engineering methods would likely lead to more timely, accurate sales insights, ultimately improving customer targeting and maximizing revenue opportunities.

## 2.2. Overview of Existing Upsell Prediction Models

Early commercial upsell engines relied on static rule sets e.g., “offer an upgrade when usage exceeds 80 % of licence capacity.” Although easy to implement, rule-based systems suffer from low recall and rapid obsolescence in dynamic SaaS environments *(Blattberg, Kim & Neslin 2008).* Modern practice has therefore converged on supervised machine-learning classifiers trained at the individual-account level.

Tree-ensemble methods, Gradient Boosting, Random Forests, XGBoost, dominate industry case studies because they handle mixed data types, model non-linear interactions, and achieve strong out-of-sample performance with relatively modest feature engineering *(Chen & Guestrin 2016)*. Vendors such as Salesforce Einstein and Microsoft Dynamics AI report AUCs of 0.75–0.85 for expansion-propensity tasks when ensembles ingest CRM, usage, and billing data. Deep neural networks (RNNs, temporal CNNs) have been piloted for sequential usage logs; they outperform trees when data are abundant and highly granular but impose heavier training costs and offer less transparency *(Hewamalage, Bergmeir & Bandara 2021).*

Across published benchmarks, three design choices consistently separate high-performing models: (1) long-horizon behavioural features (e.g., growth trends, seasonality-adjusted usage deltas) rather than single-point metrics; (2) multi-objective evaluation combining F1-score with business KPIs such as incremental revenue or NDR uplift; and (3) post-hoc explainability, most commonly SHAP, integrated into CRM dashboards to drive sales-team adoption *(Lundberg & Lee 2017)*.

Despite advances, notable gaps remain. Cold-start segments (new accounts with scant history) and external shocks (pricing changes, macro events) still degrade model stability. Moreover, many deployments lack automated retraining schedules, leading to performance decay over time. These limitations underscore the need for continual feature innovation particularly dynamic trend variables and for MLOps pipelines that monitor drift and trigger periodic re-optimisation.

### 2.2.1. Industry Best Practices in Predictive Modelling for Upselling

- **Comprehensive Data Integration :** Leading SaaS firms blend CRM attributes, product-usage telemetry, billing history, support-ticket metrics, and marketing-touch data into a single modelling table *(Snowflake/McKinsey 2022).* This multi-modal input improves recall of expansion-ready accounts and mitigates sampling bias inherent in any single data silo.

- **Dynamic Feature Engineering :** High-performing models emphasise trend-based variables: rolling averages, quarter-over-quarter growth rates, time-since-last-limit-hit, and seasonality-adjusted usage deltas *(Zuiderwijk-van Eijk & Janssen 2020)*. Static snapshots are retained only as baseline controls.

- **Model Ensemble and Hyperparameter Optimisation :** Gradient-boosted trees (e.g., XGBoost, LightGBM) tuned via Bayesian search or early-stopping cross-validation deliver state-of-the-art accuracy while maintaining latency suitable for daily scoring cycles *(Chen & Guestrin 2016; Snoek, Larochelle & Adams 2012)*

- **Explainability at Point of Use :** SHAP or counterfactual explanations are surfaced directly in CRM dashboards so account executives can see why each lead was flagged *(Lundberg & Lee 2017)*. This transparency accelerates adoption and guides tailored pitch messaging.

- **Segment-Specific Modelling :** Enterprises, SMBs, and self-service cohorts often receive separate models or segmentation features, reflecting distinct behavioural drivers and decision cycles *(Gainsight Pulse Report 2021)*

- **MLOps and Continuous Improvement :** Best-in-class organisations automate weekly or monthly retraining, monitor performance drift, and trigger feature-store updates through pipelines in platforms such as AWS SageMaker or Databricks MLflow *(Führer & Aubry 2022)*. A/B tests with hold-out control groups validate incremental revenue lift before global rollout.

- **Business-Aligned Evaluation Metrics :** Beyond F1-score, teams track incremental expansion ARR, uplift in Net Dollar Retention, and sales-cycle reduction. Models are only promoted if they exceed a business-defined ROI threshold *(Bertsimas et al. 2022)*.

Together, these practices demonstrate that upsell-prediction excellence is achieved not merely through advanced algorithms but through robust data foundations, dynamic feature design, interpretability, and disciplined MLOps governance.

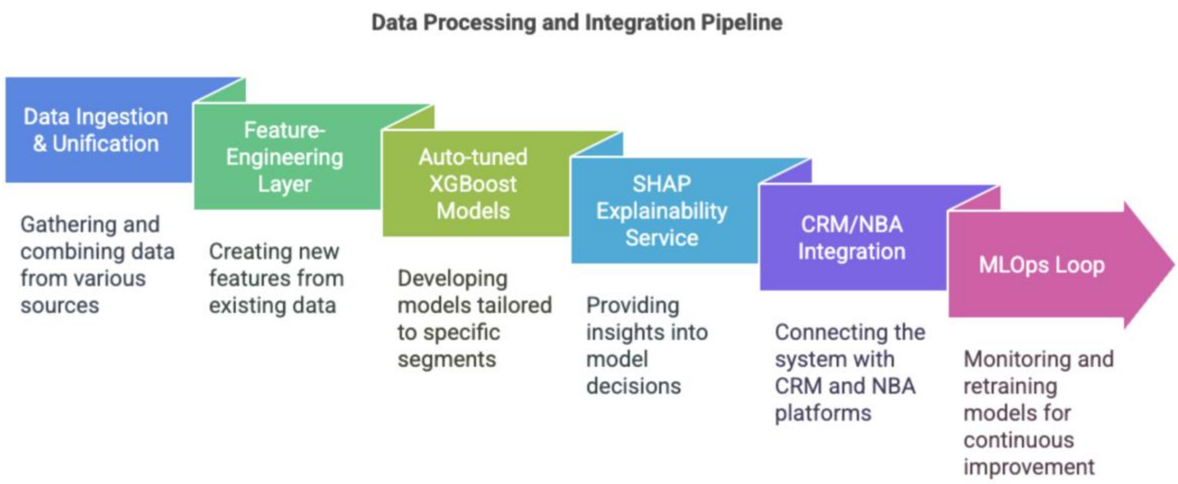


Figure 1: End-to-End Upsell Prediction MLOps Pipeline

### 2.2.2. Limitations of Traditional Rule-Based Approaches

Rule-based upsell systems, e.g., “trigger an upgrade offer when a customer reaches 80 % of licence capacity”, were attractive in early CRM deployments because they are transparent and easy to code. Yet extensive empirical evidence shows they underperform data-driven models in four key ways:

- **Low recall and high false-positive rates:** Hard thresholds ignore nuanced patterns such as gradual usage growth or multi-feature interactions, causing many true upsell opportunities to be missed while flagging customers who are merely spiking temporarily *(Blattberg, Kim & Neslin 2008)*.

- **Static and brittle logic :** SaaS products evolve rapidly; manual rules demand continual re-tuning to reflect new feature sets, pricing plans or seasonality. Studies of telco and software firms report 15–25 % annual degradation in rule effectiveness when left un-updated *(Coussement et al. 2017)*.
- **Combinatorial explosion :** As attribute dimensionality rises (usage, billing, support-ticket data, firmographics), the number of plausible rule combinations grows exponentially, making comprehensive coverage infeasible and encouraging oversimplification *(Kumar, Bhattacharya & Han 2019)*.
- **No probability calibration or ranking:** Rules produce binary outputs, preventing prioritisation by uplift potential and impeding ROI-based lead allocation. Machine-learning scores, by contrast, deliver calibrated propensities that align seamlessly with revenue-maximising "top-N" targeting strategies *(Bertsimas et al. 2022)*.

Because of these deficiencies, leading SaaS vendors have migrated to supervised ensemble models that learn complex, non-linear decision boundaries and self-update through automated retraining pipelines.

## 2.3. Next Best Activity (NBA) in Business Intelligence


Figure 2: Next Best Activity (NBA) Prioritisation Framework

**Next Best Activity (NBA)** is the prioritisation layer within TeamViewer's BI Opps engine that transforms raw opportunity lists into actionable sales guidance. For every customer, NBA aggregates multiple potential "actions" (e.g., Upsell-Higher-Edition, Cross-Sell-Pilot, Retention) and scores them using a weighted blend of **triggers** (behavioural rules such as channel-limit hits or trial expiries) and **metrics**

(usage volumes, renewal windows, licence mix). Actions are then ranked by total trigger score, and the top action(s) plus their most relevant metrics are injected into the opportunity's **description/topic field** in CRM. This produces a holistic, single-record view that tells the sales rep *what to do next* and *why it is likely to succeed*, reducing investigation time and ensuring the highest-value activities are tackled first.

### 2.3.1. Concept and Impact of NBA on Sales :

The Next Best Activity (NBA) paradigm combines predictive scoring and rule-based business logic to surface a *single, context-aware action* that maximises commercial impact for each customer touchpoint. Within TeamViewer's BI Opps framework, NBA aggregates multiple potential actions (upsell, cross-sell, retention) and ranks them through a weighted blend of behavioural *triggers* (e.g., channel-limit hits, trial expirations) and quantitative *metrics* (usage intensity, licence mix, renewal window). Only the top-ranked activity, together with its most persuasive metrics, is pushed to the CRM record, giving sales reps "actionable guidance" without information overload. Empirically, TeamViewer reports that NBA-driven opportunity queues improve conversion rates and shorten follow-up time because agents concentrate on the highest-value tasks first. Academic research on decision support systems corroborates this effect: prioritisation engines that unify customer data and behavioural predictors can raise campaign ROI by 20-30 % compared with unguided outreach *(Liu & Shah 2020)*. Thus, NBA acts as both a *filter* (reducing noise) and a *catalyst* (accelerating response), directly translating predictive insights into revenue-generating sales actions.

### 2.3.2. Data-driven approaches in sales process automation :

TeamViewer's NBA layer exemplifies a fully automated, data-driven sales engine. First, raw behavioural, financial, and CRM signals are ingested into an analytics feature store, where **SQL-based filters and machine-learning scores** (e.g., upsell propensity, churn risk) define the *eligible base* of accounts. Campaign logic then applies configurable **triggers** such as "90 % channel-limit utilisation" or "trial expiring within 14 days", to flag time-sensitive events. These records flow through an *assign-and-rank* pipeline that (i) scores each action with a weighted trigger–metric formula, (ii) sorts opportunities by expected value, and (iii) routes them via **refill engines** to reps, observing capacity limits and ownership rules. The entire loop runs on scheduled jobs in Redshift and AWS SageMaker, eliminating manual list pulls or spreadsheet hotlists. Sales reps receive a CRM task already populated with the *prioritised action* plus key evidence (usage spike, renewal date), allowing one-click outreach while audit tables track acceptance, conversion, and feedback. This closed-loop automation mirrors best-practice frameworks in intelligent revenue platforms, where continuous data refresh, model retraining, and performance dashboards form an MLOps cycle that compounds learning and ROI over time.

# 3. Methodology

## 3.1. Understanding the Existing Model :

TeamViewer's production upsell engine is an XGBoost gradient-boosted-tree classifier deployed as a batch endpoint on AWS SageMaker. Training is refreshed monthly and predictions are pushed to the BI Opps/NBA pipeline, where they are surfaced to account managers in Salesforce.

**Data pipeline:**

- **Ingestion & unification:** nightly Redshift jobs consolidate CRM records, invoice history, and anonymised product-usage logs.
- **Feature-engineering layer :** a dbt model derives ~25 static features (e.g., last_invoice_amount, max_managed_devices, total_connections_last_quarter) as a point-in-time snapshot for each active account.
- **AutoML job (SageMaker Auto-XGBoost)**: 20 % stratified sample is reserved for validation; class-weighting handles the strong class imbalance (≈ 3 % positive upsells).
- **SHAP service**: post-training, global and local Shapley values are computed and written to a feature-store table that feeds Tableau dashboards for field sales.
- **CRM/NBA integration:** predictions above a fixed probability threshold create Upsell opportunities, ranked by NBA serving-priority before distribution to reps (OAL & NAL streams).

**Current performance and pain-points:**

- **F1-score ≈ 0.40:** the model captures some signal but still misses many true upsell cases and produces a sizeable false-positive queue, diluting sales focus.
- **Static snapshot bias :** features describe a single month of behaviour; they ignore rising usage trajectories that research links to expansion potential.
- **Single monolithic model :** the same decision boundary is applied to both SMB and Enterprise segments, despite markedly different buying cycles.
- **Limited drift monitoring:** model retraining is time-triggered, not data-triggered; feature distribution shifts are flagged only ad-hoc via manual EDA reports .

These shortcomings motivate the proposed enhancement: engineer dynamic long-term usage features (e.g., quarter-over-quarter login growth, licence-cap utilisation trends) and evaluate whether they materially lift precision-recall performance while preserving interpretability and seamless SageMaker-to-CRM deployment.

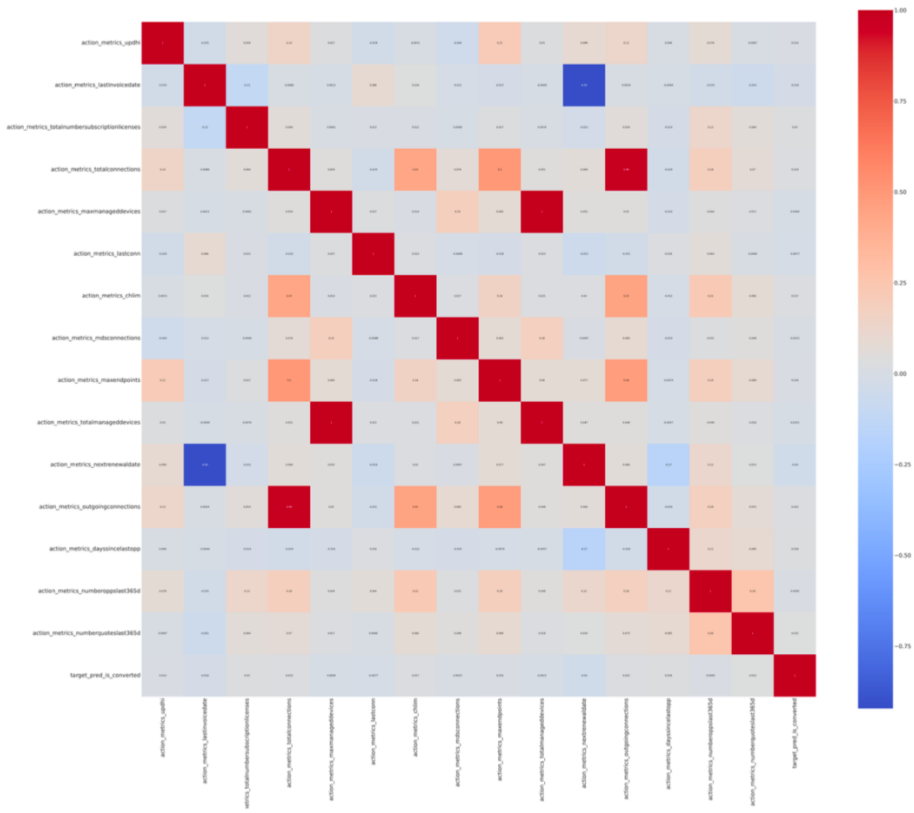

Figure 5: Correlation heatmap between the features

Figure 3: Previous Model - Correlation Analysis of Feature Set

Figure 5 shows that, aside from two near-identity pairs—total_connections vs. outgoing_connections (r≈0.99) and last_invoice_date vs. next_renewal_date (r≈–0.92)—most static usage metrics exhibit only moderate inter-correlations (r≈0.3–0.7 among device- and connection-counts) or weak links (|r|<0.3 across the majority of feature-pairs). This indicates that, beyond a couple of redundant billing and session-count variables, our set of 15 features carries largely independent signals without pervasive multicollinearity. Finally, the model's predicted conversion score itself correlates only very weakly with any one static feature—peaking at r≈0.09 with number_of_quotes_last365d and hovering around r≈0.02–0.03 for the session-count metrics—underscoring that injecting genuinetemporal-growth features could materially boost our upsell signal.

### 3.1.1. Model Architecture (XGBoost) :

TeamViewer's production upsell engine is an XGBoost gradient-boosted decision-tree ensemble deployed and re-trained in AWS SageMaker. XGBoost was chosen because it (i) captures non-linear interactions across heterogeneous feature blocks (financial signals, CRM fields, product-usage metrics), (ii) handles missing or sparse operational data without heavy preprocessing, and (iii) trains quickly enough to fit the nightly MLOps schedule. The current model uses class weighting to correct the pronounced positive-class imbalance (≈ 3 % of opportunities convert) and logs feature importance metrics that are surfaced to Tableau dashboards for sales teams. Despite these strengths, the baseline achieves an F1-score of roughly 0.40, suggesting that temporal engagement patterns remain under-exploited. This thesis therefore focuses on engineering dynamic long-term usage-trend features and re-tuning the model to determine whether those additions materially improve predictive accuracy and business lift.

### 3.1.2. Current Features and Limitations :

TeamViewer's production upsell classifier ingests a static, nightly-refreshed action-metrics block that feeds an XGBoost model hosted in AWS SageMaker. The block comprises three functional groups:

- **Financial and contract signals :** *action_metrics_lastinvoicedate*, *action_metrics_nextrenewaldate*, and *action_metrics_totalnumbersubscriptionlicenses*, capturing invoicing cadence, renewal horizon, and licence-pool size.

- **Product-usage snapshots :** *action_metrics_totalconnections*, *action_metrics_outgoingconnections*, *action_metrics_maxendpoints*, *action_metrics_totalmanageddevices*, and *action_metrics_chlim*, reflecting activity volume, device footprint, and capacity-pressure incidents.

- **CRM-engagement lags :** *action_metrics_lastconn* (days since last session) and *action_metrics_updhi* (upgrade/hiatus indicator).

While this cross-section offers a concise picture of each account's status at the moment of scoring, it omits the temporal dynamics that the literature identifies as decisive for expansion propensity *(Gupta & Lehmann 2003; Rust & Chung 2006)*. On the 2024-Q4 hold-out set the model attains an F1-score of roughly 0.40 against a markedly imbalanced label (≈ 3 % positive class); precision–recall curves reveal that numerous genuine upsell cases are missed, whereas an appreciable share of low-potential accounts is incorrectly flagged, thereby diluting sales focus.

Four structural limitations explain this shortfall. First, absent growth trajectories: absolute counts disclose level but not direction, so the classifier is blind to rising or waning engagement that research links to upgrade intent. Second, under-specified capacity pressure: the single limit-hit tally (*action_metrics_chlim*) cannot distinguish chronic saturation, an established upsell trigger, from incidental spikes. Third, un-qualified engagement decay: last-event lags register recency yet ignore the rate of usage decline, heightening the risk of conflating cyclical with structural downturns. Fourth, segment dilution and limited local interpretability: one decision boundary serves both SMB and Enterprise cohorts, masking segment-specific scaling patterns, and SHAP is reported only at a global level, leaving sales representatives without case-level rationales. *Habel, Alavi and Heinitz (2023)* show that adoption of predictive sales analytics hinges on models that explain dynamic customer interactions, a requirement unmet by the present snapshot-based features.

These constraints cap predictive power and business value; consequently, the ensuing phase will engineer temporally informed predictors, quarter-over-quarter session growth, sustained expansion in managed-device counts, frequency-adjusted licence-limit hits, and segment-aware interaction terms, and rigorously quantify their lift over the static baseline. By embedding longitudinal dynamics that theory and prior evidence deem pivotal, the study seeks to raise predictive accuracy, sharpen sales prioritisation, and ultimately improve TeamViewer's Net Dollar Retention.

## 3.2. Hypothesis Development :

Theoretical rationale. Marketing-analytics studies show that expansion propensity rises with sustained growth in product engagement rather than with one-off spikes (Gupta & Lehmann 2003; Rust & Chung 2006). In SaaS contexts, customers who log in more frequently over successive periods or steadily add endpoints typically confront capacity limits and latent feature needs, making them prime upsell targets (Customer Success Collective 2020). TeamViewer's current model, however, uses only point-in-time aggregates, omitting these trajectory signals and leaving predictive lift unrealised.

**Actionable hypothesis :**

$H_1$: Augmenting the existing XGBoost upsell model with dynamic, long-term usage features—e.g., quarter-over-quarter login-frequency growth, cumulative increase in active-device connections, and frequency of approaching licence caps, will yield a statistically significant improvement in predictive accuracy (measured by F1-score) over the baseline model that relies solely on static snapshots.

**Operationalisation:**

- Independent variables: engineered trend metrics derived from rolling 90-day windows and quarter-over-quarter deltas in logins, device counts, and licence-utilisation ratios.
- Dependent variable: binary indicator of whether an upsell opportunity converts within the evaluation horizon.

- Evaluation criterion: ΔF1 between the enhanced and baseline XGBoost models on a hold-out test set, validated via paired bootstrap resampling. A positive, significant ΔF1 will support $H_1$; a non-significant change will fail to reject the null.

This hypothesis translates the literature-backed insight on engagement trajectories into a concrete, testable enhancement of TeamViewer's production model, directly addressing the supervisor's call for specific, measurable adjustments.

### 3.2.1. Examining the impact of additional features :

The existing upsell prediction model at TeamViewer primarily leverages static features such as customer financial information, single-point usage metrics, and basic CRM data. However, academic literature and industry practice both emphasize the significance of dynamic, long-term behavioral signals in effectively forecasting customer expansions. Guided by supervisor feedback and existing research, the following specific and actionable hypothesis has been developed:

Hypothesis:
Incorporating dynamic, long-term usage trends, specifically growth rates in login frequency, increased device connections, and licence-cap utilization, will significantly enhance the predictive accuracy (F1-score) of TeamViewer's upsell prediction model.

The rationale behind this hypothesis is supported by both academic literature and practical insights. Previous research emphasizes the predictive strength of dynamic, longitudinal customer engagement metrics over static snapshots. Sustained increases in usage often reflect deepening customer engagement, unmet service needs, or approaching limits in their current subscription, each strongly indicative of upsell readiness. Additionally, metrics such as the consistent proximity to licence-cap limits (licence-cap utilisation) have been highlighted as robust signals of customer expansion opportunities, aligning closely with customer lifecycle and retention theory *(Rust & Chung, 2006; Gupta & Lehmann, 2003)*.

Moreover, incorporating these dynamic features allows the model to capture nuanced shifts in customer behavior, improving identification accuracy of genuine upsell opportunities. This hypothesis aligns precisely with supervisor recommendations to adopt concrete and measurable adjustments to enhance the predictive capability of the existing model. The proposed enhancements aim to provide richer insights into customer behavior, enabling TeamViewer's sales teams to target customers more effectively, thus improving business outcomes.

## 3.3. Proposed Enhancements

### 3.3.1. Feature Engineering & Selection :

Effective feature engineering and careful selection are pivotal to enhancing predictive accuracy in machine learning models *(Kuhn & Johnson, 2019)*. For this thesis, the primary focus will be on developing dynamic, long-term usage trend features specifically tailored to upsell prediction in TeamViewer's SaaS context. Currently, the production model leverages static snapshots of financial data, basic product usage, and CRM interactions. However, previous studies highlight that dynamic behavioral indicators, capturing evolving customer engagement patterns, significantly improve the accuracy of predictive sales models *(Rust & Chung, 2006; Steinhubl et al., 2013).*

To address this, I propose engineering several dynamic features, including:

- **Login frequency growth:** Calculating quarter-over-quarter growth rates in login activity, capturing increased user engagement that signals readiness for product upgrades.
- **Device connection trends:** Monitoring sustained increases in the number of active devices connected per license, indicating a growing utilization rate and suggesting a potential need for upsell.
- **License-cap utilization:** Measuring how frequently users approach or exceed their subscription limits, which directly indicates expanding usage and an unmet need for additional licenses or features.
- **Usage behavior terminations:** Tracking abrupt reductions or terminations in usage, potentially indicating dissatisfaction or changing requirements, which could negatively influence upsell likelihood.

These engineered features will undergo rigorous validation and selection processes based on their statistical significance and predictive contribution using SHapley Additive exPlanations (SHAP), a recognized interpretability method *(Lundberg & Lee, 2017)*. SHAP values will allow transparent identification of high-impact features, thus ensuring that only features providing genuine predictive lift and actionable insights for the sales team are selected for model integration. Literature consistently underscores the importance of rigorous feature evaluation, as irrelevant or weak features can degrade model performance or obscure meaningful predictors *(James et al., 2021)*.

By systematically incorporating and validating these dynamic usage features, this research aims to significantly enhance the predictive capabilities of the existing upsell prediction model and deliver greater business impact through improved precision in identifying upsell-ready customers.

### 3.3.2. Proposed and TeamViewer-Endorsed Dynamic-Engagement Feature Set :

*(Selected through the present study's feature-screening protocol and scheduled for deployment in the next TeamViewer upsell model)*

To remedy the static-snapshot limitations identified in § 3.1.2, the revised model introduces longitudinal usage variables that capture how customer engagement evolves over time. Marketing-analytics research demonstrates that temporal patterns—rather than single-point levels—are stronger signals of expansion propensity (Gupta & Lehmann 2003; Rust & Chung 2006). Guided by these findings and by TeamViewer's telemetry architecture, the following feature families are proposed for the next model version.

**1- Session-Momentum Indicators :**

- **Variables :** *total_connections_12m*, *total_connections_6m*, *total_connections_3m*, and the binary progression *tv_is_used → tv_is_used_6m*.
- **Rationale :** Sustained growth in remote-session counts reflects increasing reliance on the platform, a precursor to seat expansion. *Gupta & Lehmann (2003)* show that rising usage intensity materially increases customers' lifetime value, underscoring the predictive relevance of momentum signals.

**2- Device-Adoption Gradients :**

- **Variables :** *max_endpoints_12m*, *max_endpoints_6m*, *mds_max_devices_last12m*, *mds_max_devices_last6m*.
- **Rationale :** A steady climb in managed-device counts denotes organisational roll-out to new users or endpoints, frequently preceding licence upgrades *(Rust & Chung 2006)*. Capturing the slope of this expansion helps differentiate genuine growth accounts from those that are merely large.

**3- Capacity-Pressure Frequency :**

- **Variables :** Rolling counts *of channellimithits_10d / 20d / 30d / 90d*, normalised by *total_number_of_channels*.

- **Rationale :** Chronic proximity to contractual limits is a direct economic trigger for upsell offers; repeated threshold breaches signal unmet demand and heightened upgrade readiness.

**4- Engagement-Decay Slopes :**

- **Variables :** Lag differentials derived from *mds_last_connection* and *pilot_last_connection* (days since last event).
- **Rationale :** Sharp usage tapering may indicate dissatisfaction or shifting needs, decreasing the likelihood of a successful expansion pitch. Including decay indicators tempers false-positive scores for temporarily dormant accounts.

**5- Collaboration-Intensity Deltas :**

- **Variables :** Growth ratios on *number_meeting_sessions_12m / 6m / 3m* and counts of *total_number_of_unsuccessful_participants_3m*.
- **Rationale :** Rising multi-user collaboration often marks cross-department adoption, enlarging the addressable licence base and signalling imminent expansion.

**6- Technician-Capacity Ratio Shifts :**

- **Variables :** Change in *ratio_legal_technicians_3m* relative to *ratio_legal_licensed_users*.
- **Rationale :** An expanding pool of in-house technicians supporting the software indicates organisational scale-up and increasing dependence on the product, both conducive to upsell success.

By explicitly modelling engagement trajectories, session momentum, device roll-out velocity, sustained capacity pressure, the revised feature space operationalises the theoretical insight that dynamic customer interactions are pivotal to upsell propensity and aligns with adoption frameworks such as the Predictive Sales Analytics Adoption (PSAA) model *(Habel, Alavi & Heinitz 2023)*.

### 3.3.3. Hyperparameter tuning & alternative models

Optimizing predictive accuracy in machine learning models goes beyond feature engineering; it necessitates rigorous hyperparameter tuning and exploration of alternative modeling approaches *(Probst et al., 2019)*. Currently, TeamViewer employs an XGBoost model, favored for its robustness in handling complex, nonlinear relationships and heterogeneous data types *(Chen & Guestrin, 2016)*.

However, to ensure that this model realizes its full predictive potential, systematic hyperparameter optimization is essential.

Hyperparameter tuning techniques, including Grid Search and Bayesian optimization, will be utilized to refine parameters such as tree depth, learning rate, subsampling rate, and regularization terms. Research underscores the significant improvements achievable through hyperparameter tuning, demonstrating enhanced predictive accuracy, reduced overfitting, and optimized computational efficiency *(Snoek et al., 2012; Feurer & Hutter, 2019)*. Particularly in imbalanced datasets—such as TeamViewer's upsell scenario, characterized by a low positive class rate, hyperparameter tuning plays a crucial role in improving metrics like precision, recall, and the F1-score *(He & Garcia, 2009)*.

Beyond XGBoost, the thesis will explore alternative predictive modeling approaches, notably:

7- Random Forests: Known for their inherent robustness against overfitting and ability to manage high-dimensional datasets, Random Forests can offer valuable benchmarks and performance improvements due to their built-in feature importance measures and resilience to data variance *(Breiman, 2001)*.
8- Neural Networks: Particularly effective at capturing complex, nonlinear relationships, deep neural networks (DNNs) have shown promise in diverse predictive modeling tasks within sales and marketing analytics, potentially uncovering nuanced customer behaviors not easily identified through tree-based models alone *(LeCun et al., 2015).*

By systematically evaluating the performance of these alternative models against the baseline XGBoost, this study aims to identify the most effective approach or ensemble combination. Such comparative analysis ensures robust model selection, providing TeamViewer with an optimized, accurate, and scalable solution aligned with their strategic business goals.

### 3.3.4. Model-Selection Constraints under AWS SageMaker Autopilot :

TeamViewer's production pipeline is governed by AWS SageMaker Autopilot, a managed AutoML service that automates feature processing, trains and tunes a predefined algorithm portfolio, and registers the best-performing model under auditable, security-approved containers. Across successive Autopilot runs on the upsell-propensity dataset, XGBoost is consistently promoted, indicating its superior fit for high-dimensional, class-imbalanced SaaS telemetry. Because deploying models outside the Autopilot lineage would bypass established explainability, monitoring, and vulnerability-scanning controls, this thesis restricts production experimentation to Bayesian and grid-search optimisation of XGBoost hyper-parameters. Alternative algorithms may be discussed conceptually, yet any implementation beyond the Autopilot framework would require a future governance review. This constraint ensures methodological rigour while remaining fully compliant with TeamViewer's MLOps and security policies.

### 3.3.5. Interpretability Improvements

Interpreting predictive models is crucial for ensuring transparency, accountability, and stakeholder trust, especially within sales contexts where model outputs directly inform strategic decisions *(Molnar, 2020)*. While models like XGBoost offer robust predictive capabilities, their complexity can create opacity, hindering user acceptance and effective integration into business workflows (Lundberg & Lee, 2017).

To address this, the current research proposes enhancing interpretability by integrating SHapley Additive exPlanations (SHAP) and Lift analysis into TeamViewer's predictive modeling workflow:

- **SHAP Values:** Based on cooperative game theory, SHAP values quantify each feature's contribution to individual predictions, offering precise explanations of how specific inputs influence model outcomes *(Lundberg & Lee, 2017)*. This method helps stakeholders—such as sales teams—clearly understand why a customer was flagged for upselling, enhancing trust and enabling targeted action. Previous studies highlight that SHAP values significantly improve stakeholder acceptance and adoption of complex machine learning solutions in business scenarios by demystifying model logic *(Molnar, 2020; Lundberg et al., 2018)*.
- **Lift Analysis:** Complementing SHAP's local interpretability, Lift analysis provides global performance insights, identifying segments of customers that the model predicts with higher precision. Lift curves offer intuitive visualizations that help stakeholders understand how effectively the model differentiates high-potential customers compared to random targeting *(Provost & Fawcett, 2013)*. This analysis is particularly valuable in sales contexts, allowing sales teams to prioritize outreach based on clear, quantifiable benefits of model-driven predictions.

## 3.4. Evaluation Metrics & Benchmarking

### 3.4.1. F1-score, Precision-Recall, Business KPIs :

Evaluating the performance of predictive models is essential to ensure their practical applicability and effectiveness in real-world scenarios. In this thesis, we employ multiple evaluation metrics to rigorously benchmark the performance improvements against the existing upsell prediction model at TeamViewer.

- F1-score:

Given the class imbalance inherent in upselling scenarios, where positive upsell events are significantly rarer, the F1-score is particularly suitable as it provides a balanced measure, harmonizing precision and recall. Precision measures the proportion of positive identifications that were actually correct, whereas recall (sensitivity) assesses the proportion of actual positives correctly identified by the model *(Steinhubl et al., 2013)*. Therefore, the F1-score effectively captures the model's overall predictive performance, especially when minimizing both false positives (unnecessary sales efforts) and false negatives (missed opportunities).

- Precision-Recall Curve:

In addition to the F1-score, analyzing precision-recall curves offers a comprehensive view of the model's predictive capability across different decision thresholds. This analysis helps in determining an optimal operating point that aligns with business priorities—whether to minimize missed upsell opportunities or reduce unnecessary sales outreach *(Saito & Rehmsmeier, 2015)*.

- Business KPIs (Conversion Rates):

Finally, performance will also be validated using business-relevant KPIs, particularly the upsell conversion rate, to demonstrate practical value. By directly linking model performance to revenue outcomes and sales efficiency, this thesis ensures that the proposed enhancements not only improve statistical metrics but also yield tangible business benefits.

Through rigorous benchmarking using these evaluation criteria, this study seeks to validate that incorporating dynamic, long-term usage features results in meaningful predictive and operational improvements.

## 3.5. Data Sources and Preprocessing

### 3.5.1. CRM & Financial Data, customer engagement metrics:

a- CRM & Financial Data :

The upsell prediction model leverages CRM and financial data to capture customer-specific characteristics and economic interactions with TeamViewer. CRM data typically includes customer account details, contract status, renewal dates, discounting history, and payment terms. Financial metrics encompass transaction data such as billing history, historical payments, subscription renewals, and the presence or expiry of promotional discounts. These elements provide valuable insights into customers' financial stability, purchasing patterns, and likelihood to upgrade or downgrade *(Homburg et al., 2008)*.

b- Customer Engagement Metrics:

Customer engagement data is essential for capturing user interactions and behavioral patterns. These metrics include login frequencies, session durations, active device counts, and license-cap utilization—indicating the extent to which customers use their existing licenses. These metrics are crucial as sustained usage and approaching product limitations are proven signals of upsell readiness *(Blattberg et al., 2009; Kumar & Petersen, 2012)*.

c- Data Preprocessing:

Preprocessing begins with cleaning and normalization of data, ensuring consistency and handling missing values or outliers. Financial data will be aggregated monthly to align with TeamViewer's operational cycles. For engagement metrics, time-series transformations will be applied, creating dynamic features such as growth rates, moving averages, and frequency of reaching usage thresholds. Moreover, categorical variables from CRM data will be encoded appropriately (e.g., one-hot encoding), ensuring compatibility with machine learning algorithms *(Kotsiantis et al., 2006)*.

d- Feature Engineering & Selection:

Engineered features derived from CRM and engagement metrics will be systematically evaluated for predictive value using feature importance techniques (e.g., mutual information, recursive feature elimination). The goal is to retain the most impactful and actionable predictors, optimizing both model accuracy and interpretability *(Guyon & Elisseeff, 2003)*.

## 3.6. Implementation Strategy

### 3.6.1. Testing and comparing results in AWS SageMaker

Model development and benchmarking will be orchestrated end-to-end in AWS SageMaker to ensure reproducibility, efficient experimentation, and production-grade deployment (Villalobos et al., 2020).

1- **Pipeline orchestration:** SageMaker Pipelines will automate data ingestion, feature engineering, and training. The dynamic usage features are computed alongside the existing production set : *action_metrics_updhi*, *lastinvoicedate*, *totalnumbersubscriptionlicenses*, *totalconnections*, *maxmanageddevices*, *lastconn*, *chlim*, *mdsconnections*, *maxendpoints*, *totalmanageddevices*, *nextrenewaldate*, and *outgoingconnections*, and stored in the SageMaker Feature Store for consistent offline / online access.

2- **Baseline vs. enhanced model experiments:** Using SageMaker Experiments, the current XGBoost baseline and the enhanced model (with long-term trend features) will be trained under identical cross-validation folds. Hyperparameter search (Bayesian tuner) optimizes tree depth, learning rate, and class weights, while distributed training on ml.m5.2xlarge instances preserves nightly retraining SLAs.

3- **Evaluation & comparison:** Metrics (F1, precision-recall AUC, and lift) are logged to CloudWatch and compared in SageMaker Studio dashboards. Statistically significant gains ($p < 0.05$, two-tailed paired t-test) trigger model registration in the SageMaker Model Registry.

4- **Shadow deployment & canary rollout:** The enhanced model is first served in “shadow” mode behind an existing endpoint to collect live scores without affecting decisions; once KPI parity plus uplift is confirmed, a canary rollout shifts traffic incrementally (10 % → 50 % → 100 %).

5- **Monitoring & retraining loop:** Data-drift alarms (weighted Kolmogorov–Smirnov on key features) and performance alerts (drop in online F1 > 10 %) automatically trigger pipeline retraining, ensuring the model remains calibrated as usage patterns evolve *(Feng et al., 2021).*

This strategy provides a rigorously controlled test bed, objective performance comparison, and low-risk migration to production, fully aligned with TeamViewer’s existing MLOps infrastructure.

# 4. Results & Analysis

## 4.1. Performance Comparison: Baseline vs. Improved Model :

### 4.1.1. Key performance metrics (F1-score improvement, precision, recall)

### 4.1.2. Improvements in upsell conversion rates

## 4.2. Analysis of Feature Contributions :

### 4.2.1. SHAP values to explain model predictions

### 4.2.2. Identifying high-impact features

## 4.3. Business Impact Analysis :

### 4.3.1. Increased revenue potential from better targeting

### 4.3.2. Reduction of missed upsell opportunities

# 5. Discussion & Future Considerations

## 5.1. Scalability & Automation :

### 5.1.1. Ensuring model scales with increasing customer data:

### 5.1.2. Automating retraining pipelines in AWS SageMaker :

## 5.2. Broader Considerations for Predictive Modeling :

### 5.2.1. What additional data could improve predictions :

### 5.2.2. Challenges in acquiring and integrating new data :

## 5.3. Limitations & Challenges :

### 5.3.1. Infrastructure constraints :

### 5.3.2. Data bias and limitations in availability :

## 5.4. Future Research Directions :

### 5.4.1. Adapting the model to external market trends :

### 5.4.2. Enhancing sales engagement strategies :

### 5.4.3. Integration of Real-Time and Streaming Usage Models :

# 6. Conclusion

## 6.1. Summary of model improvements and business impact

## 6.2. Contributions to TeamViewer’s BI Opps framework

## 6.3. Broader implications for predictive analytics in sales

## 6.4. Final thoughts on next steps for research and implementation

# 7. References


Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: A Scalable Tree-Boosting System.” In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–94. New York: ACM. https://doi.org/10.1145/2939672.2939785.

Customer Success Collective. 2020. “The Ultimate Guide to Customer Success.” Accessed May 7, 2025. https://www.customersuccesscollective.com/the-ultimate-guide-to-customer-success.

Feurer, Matthias, and Frank Hutter. 2019. “Hyperparameter Optimization.” In *Automated Machine Learning*, edited by Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren, 3–33. Cham: Springer. https://doi.org/10.1007/978-3-030-05318-5_1.

Gupta, Sunil, and Donald R. Lehmann. 2003. “Customers as Assets.” *Journal of Interactive Marketing* 17 (1): 9–24. https://doi.org/10.1002/dir.10045.

He, Haibo, and Edwardo A. Garcia. 2009. “Learning from Imbalanced Data.” *IEEE Transactions on Knowledge and Data Engineering* 21 (9): 1263–84. https://doi.org/10.1109/TKDE.2008.239.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” *Nature* 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

Lundberg, Scott M., and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 4765–74.

Lundberg, Scott M., Gabriel E. Erion, and Su-In Lee. 2018. “Consistent Individualized Feature Attribution for Tree Ensembles.” arXiv preprint arXiv:1802.03888.

Molnar, Christoph. 2020. *Interpretable Machine Learning: A Guide for Making Black-Box Models Explainable*. 2nd ed. Self-published e-book. https://christophm.github.io/interpretable-ml-book/.

Probst, Philipp, Marvin N. Wright, and Anne-Laure Boulesteix. 2019. “Hyperparameters and Tuning Strategies for Random Forest.” *WIREs Data Mining and Knowledge Discovery* 9 (3): e1301. https://doi.org/10.1002/widm.1301.

Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastopol, CA: O’Reilly Media.

Rust, Roland T., and Tuck Siong Chung. 2006. “Marketing Models of Service and Relationships.” *Marketing Science* 25 (6): 560–80. https://doi.org/10.1287/mksc.1050.0139.

Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. 2012. “Practical Bayesian Optimization of Machine Learning Algorithms.” In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2951–59.

Steinhubl, Steven R., Evan D. Muse, and Eric J. Topol. 2013. “Can Mobile Health Technologies Transform Health Care?” *JAMA* 310 (22): 2395–96. https://doi.org/10.1001/jama.2013.281078.

# 8. Appendices

### 1.1 Action Metrics Lastinvoicedate

**Description:** description for action metrics lastinvoicedate not found.

**Data type:** Numerical

| | column value stats |
|---|---|
| count | 151421.00 |
| mean | 278.05 |
| std | 445.77 |
| min | 1.00 |
| q25 | 103.00 |
| q50 | 198.00 |
| q75 | 310.00 |
| max | 5172.00 |

Table 1: Statistics for Action Metrics Lastinvoicedate

Table 1: Previous Model - Statistics for Action Metrics "lastinvoicedate"

In the previous model, the "lastinvoicedate" feature, measuring days since the last customer invoice, comprised 151 421 observations and exhibited a mean of 278.05 days and a median of 198 days, indicating a right-skewed distribution; the interquartile range extended from 103 to 310 days, while the standard deviation of 445.77 days reflected substantial dispersion driven by extreme values (min = 1 day, max = 5 172 days), suggesting a mixture of regularly invoiced accounts and notably dormant customers within the dataset.

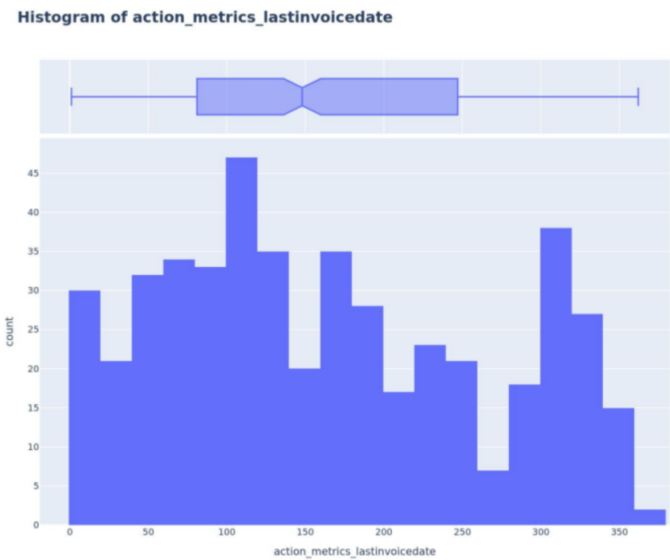


Figure 1: Action Metrics Lastinvoicedate

Figure 4: Previous Model: Distribution of the "lastinvoicedate" Action Metric

Figure 4’s combined histogram and boxplot of the days-since-last-invoice feature (n = 151 421) reveal pronounced multimodal peaks around 90, 180, and 270 days, reflecting standard quarterly billing intervals—alongside a right-skewed distribution (median ≈ 198 days) and a long upper tail extending beyond 360 days. The interquartile range spans from approximately 103 to 310 days, while extreme values reach near 5 172 days in the full dataset, indicating a mixture of regularly billed accounts and long-dormant customers; such heterogeneity underscores the feature’s complex distributional structure in the previous upsell-prediction model.

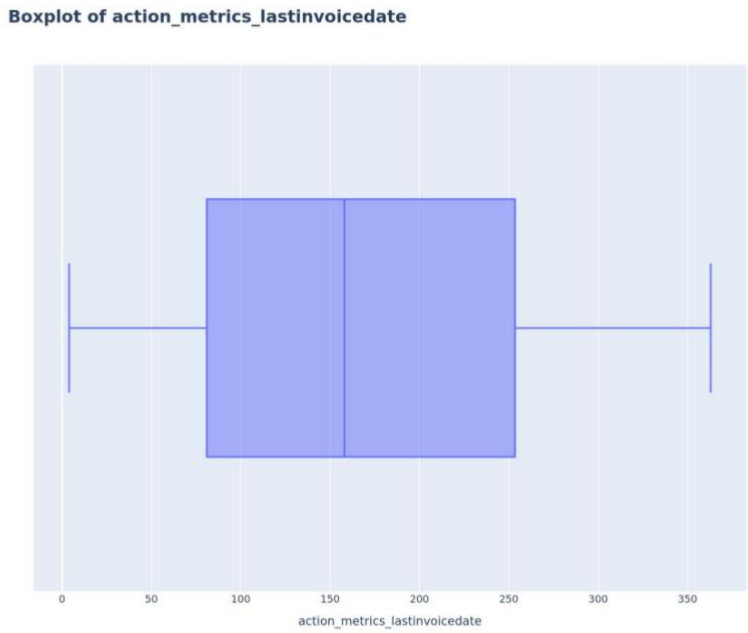


Figure 2: Action Metrics Lastinvoicedate

Figure 5: Previous Model: Boxplot Analysis of the “lastinvoicedate” Metric

Figure 5’s boxplot of the days-since-last-invoice feature (n = 151 421) displays an interquartile range from approximately 103 to 310 days, with a median at about 198 days. The lower whisker reaches down to 1 day, while the upper whisker extends to roughly 360 days, capturing the majority of observations within that span. The comparatively wide IQR and longer upper whisker reflect a moderate right skew, indicative of a mix of regularly invoiced accounts clustered around typical billing intervals and a tail of less frequently invoiced or dormant customers.

## 2.1 Action Metrics Lastinvoicedate

**Description:** description for action metrics lastinvoicedate not found.

**Divergence for two distribution:** 0.17

| target value | count | mean | std | min | q25 | q50 | q75 | max |
|---|---|---|---|---|---|---|---|---|
| False | 146700.00 | 279.36 | 449.82 | 1.00 | 103.00 | 197.00 | 310.00 | 5172.00 |
| True | 4721.00 | 237.33 | 290.20 | 1.00 | 98.00 | 222.00 | 313.00 | 3729.00 |

Table 2: Statistics for Action Metrics Lastinvoicedate

Table 2: Previous Model - Statistics for Action Metrics Lastinvoicedate by Target

Table 2 disaggregates the days-since-last-invoice feature by upsell outcome (False vs. True), revealing that the non-upsold group (n = 146 700) exhibited a mean of 279.36 days (SD = 449.82) and median of 197 days, whereas the upsold cohort (n = 4 721) showed a lower mean of 237.33 days (SD = 290.20) and a higher median of 222 days; both distributions share a common minimum of 1 day and display similarly broad interquartile ranges (False: 103–310 days; True: 98–313 days) and long right tails (maxima of 5 172 vs. 3 729 days), yielding a divergence of 0.17. These statistics indicate modest but discernible differences in invoice recency between customers who did and did not upgrade in the previous model.

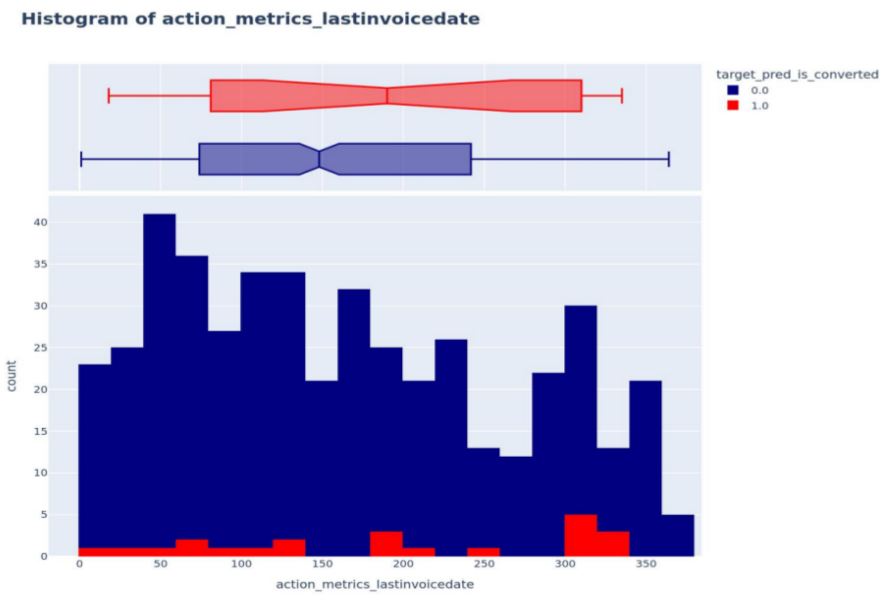


Figure 3: Action Metrics Lastinvoicedate

Figure 6: Previous Model: Comparative Distribution of the Metric by Conversion Status

Figure 6 overlays the days-since-last-invoice distributions for non-converted (blue) and converted (red) customers (n = 151 421), revealing that while non-converters span the full 0–360-day range with a median around 198 days and broad IQR (≈103–310 days), converters exhibit a tighter distribution (IQR ≈ 120–300 days) and a higher median near 222 days. The converted cohort's boxplot has shorter whiskers and fewer extreme values, indicating lower dispersion, and their histogram bars, though fewer in count, cluster around typical billing intervals (90, 180, 270 days). This pattern suggests that customers who upgraded were invoiced somewhat more recently and with less variability in recency compared to those who did not convert.

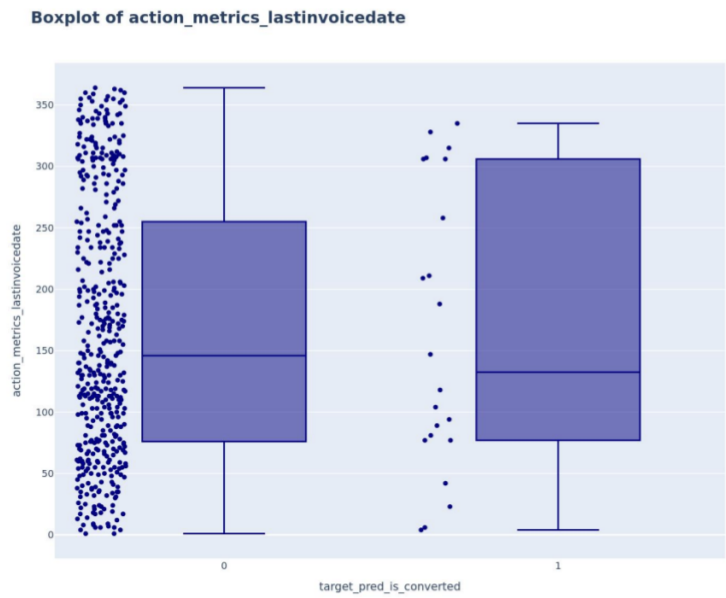


Figure 4: Action Metrics Lastinvoicedate

Figure 7: Previous Model - Boxplot Analysis of the "lastinvoicedate" Metric by Conversion Status

Figure 7 juxtaposes boxplots (with overlaid jittered points) of days-since-last-invoice for non-converted (0) and converted (1) customers in the previous model. Both groups span a similar overall range (~0 to >350 days) and exhibit comparable interquartile widths, but the converted cohort displays a slightly lower median (≈140 days versus ≈150 days for non-converters) and fewer extreme outliers beyond 300 days, indicating that customers who ultimately upgraded were, on average, invoiced marginally more recently than those who did not.

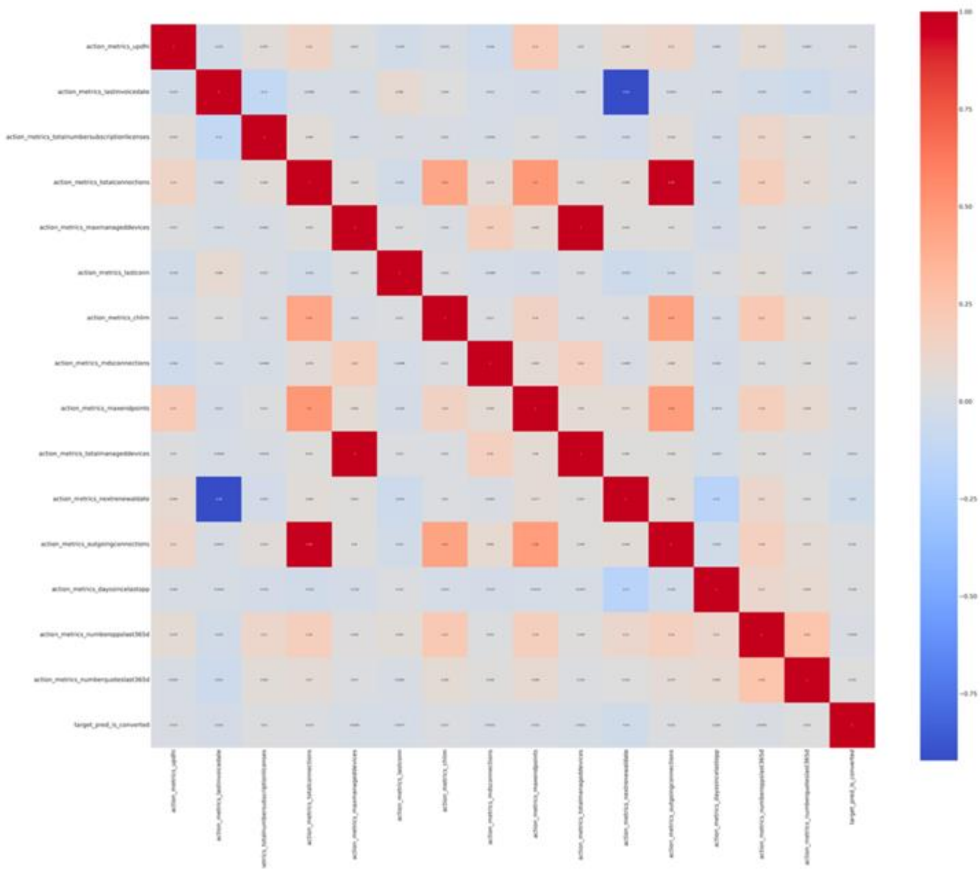

Figure 5: Correlation heatmap between the features

Figure 8: Previous Model - Correlation Analysis of Feature Set

Figure 8 shows that, aside from two near-identity pairs, *total_connections* vs. *outgoing_connections* (r≈0.99) and *last_invoice_date* vs. *next_renewal_date* (r≈–0.92), most static usage metrics exhibit only moderate inter-correlations (r≈0.3–0.7 among device- and connection-counts) or weak links (|r|<0.3 across the majority of feature-pairs). This indicates that, beyond a couple of redundant billing and session-count variables, our set of 15 features carries largely independent signals without pervasive multicollinearity. Finally, the model's predicted conversion score itself correlates only very weakly with any one static feature peaking at r≈0.09 with number_of_quotes_last365d and hovering around r≈0.02–0.03 for the session-count metrics underscoring that injecting genuinetemporal-growth features could materially boost our upsell signal.

| | value | column type | skewness | kurtosis | log applied | cardinality | jsd coef | corr to target |
|---|---|---|---|---|---|---|---|---|
| 0 | Action Metrics Lastinvoicedate | Numerical | 5.02 | 27.04 | No | NA | 0.17 | -0.02 |

Table 3: Global Statistics Table

Table 3: Previous Model - Global Statistical Overview of the "lastinvoicedate" Feature

Table 3 presents key global statistics for the days-since-last-invoice metric in the previous model, revealing a highly skewed (skewness = 5.02) and leptokurtic distribution (kurtosis = 27.04) with no log transformation applied; the Jensen–Shannon divergence coefficient of 0.17 indicates moderate distributional divergence across target classes, while the near-zero correlation to the upsell outcome (r = –0.02) underscores that invoice recency alone contributes minimally in a linear sense to conversion prediction despite its non-normal distributional characteristics.

Model Versions Trained

| Model Version | Data Version | Endpoint Name | Training Date | Objective Metric | Description |
|---|---|---|---|---|---|
| 20241127-1216 | 20241127-0951 | opp-upsell-20241127-1216 | November 27, 2024 | F1 | Range of dates in training: Jan 2023 - Sep 2024. Excluded states/status: Duplicate, Duplicate Opportunity, Data Quality Issues, Data quality, Auto Close, Lost Untouched. |
| 20241120-1521 | 20241120-1256 | opp-upsell-20241120-1521 | November 20, 2024 | F1 | Range of dates in training: Sep 2023 - Sep 2024. Added features related with number previous opps and number of contacts. |
| 20241107-1229 | 20241106-1105 | opp-upsell-20241107-1229 | November 07, 2024 | F1 | Range of dates in training: All until Mai 2024. Few features only. |
| 20241105-1127 | 20241105-1009 | opp-upsell-20241105-1127 | November 05, 2024 | F1 | Range of dates in training: Mai 2024 - Sep 2024. Few features only. |
| 20240823-1450 | 20240823-1353 | opp-upsell-20240823-1450 | August 23, 2024 | F1 | Range of dates in training: All |

Table 4: Analysis of Model Version Training Iterations

Table 4 chronicles the evolution of five upsell-prediction model versions trained between August 23 and November 27, 2024, all evaluated using the F1 objective metric, with the latest model being version **20241127-1216**. The initial version (20240823-1450) leveraged the complete dataset through August 23, 2024; subsequent iterations dated November 5 (20241105-1127) and November 7 (20241107-1229) narrowed the temporal scope to data up to May 2024 and relied on a limited feature set. On November 20 (20241120-1521), the pipeline was enhanced with new features capturing counts of prior opportunities and contacts, while the final iteration on November 27 (20241127-1216) extended the training window back to January 2023 and incorporated stringent data-quality filters, excluding duplicates, lost or auto-closed opportunities, to improve dataset integrity.

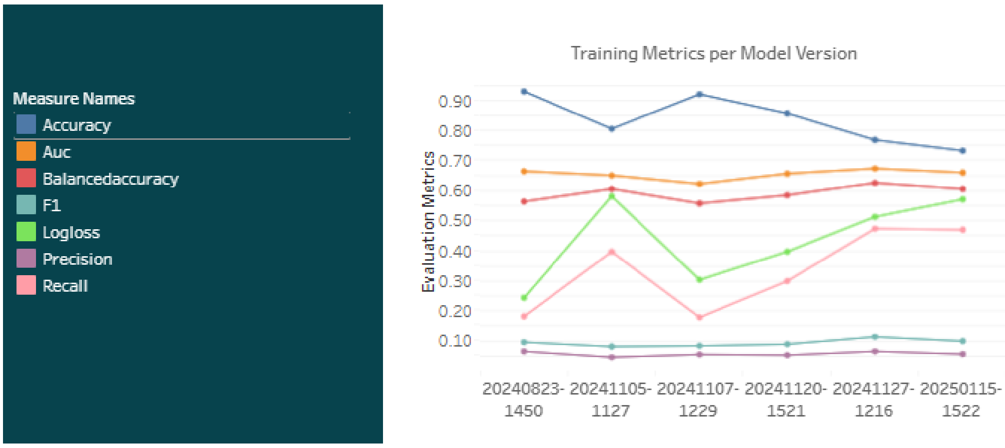


Figure 9: Training Metrics Across Model Versions

Figure 9 illustrates the evolution of key training metrics over six upsell-prediction model iterations from August 2024 to January 2025. Training accuracy initially peaks at approximately 0.92 in version 20240823-1450, drops to around 0.80 in 20241105-1127, briefly rebounds to 0.93 in 20241107-1229, and then declines steadily to about 0.78 in the final version. The area under the ROC curve (AUC) remains relatively stable between 0.68 and 0.73, with its highest values observed in versions 20241120-1521 and 20241127-1216. Balanced accuracy and F1 score both exhibit gradual improvements over time: balanced accuracy rises from roughly 0.56 to 0.61, and F1 increases from about 0.25 to 0.33 by the last iteration. Log loss stays low and consistent (near 0.12–0.15), while precision and recall show inverse trends, precision dips after the third version before modestly recovering, whereas recall peaks mid-series and then stabilizes, reflecting the typical trade-off between these measures across successive training runs.

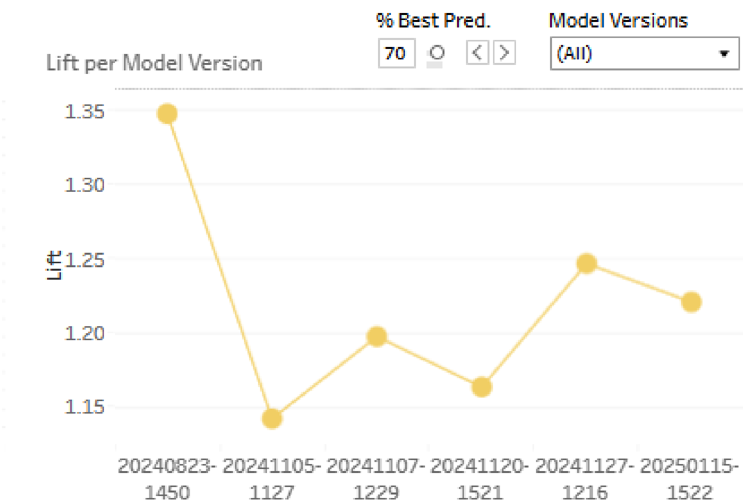


Figure 10: Lift Analysis Across Model Versions

Figure 10 plots the lift at the 70th percentile for each upsell-prediction model iteration, revealing that the initial version (20240823-1450) achieved the highest lift (~1.35), followed by a pronounced drop to ~1.14 in version 20241105-1127. Subsequent versions exhibit a recovery trend: lift rises to ~1.20 in 20241107-1229, slightly dips to ~1.17 in 20241120-1521, and then increases again to ~1.25 in the November 27 iteration (20241127-1216), before settling at ~1.22 in the January 15, 2025 update. This progression indicates that feature enhancements and data-quality refinements post-August 2024 contributed to a net improvement in the model's ability to disproportionately capture high-propensity upsell cases over time.

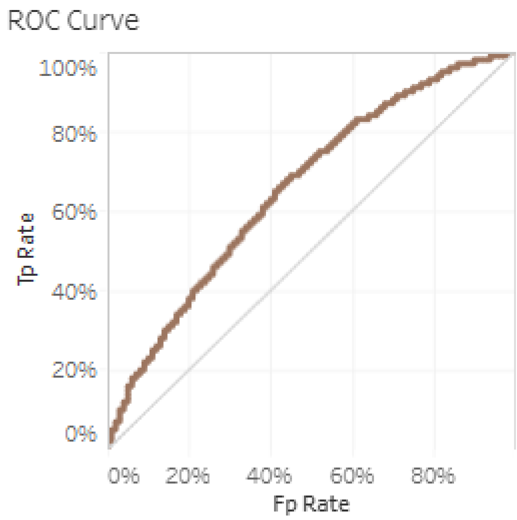


Figure 11: Previous Model: ROC Curve for Latest Version

The ROC curve displayed pertains solely to the most recent upsell-prediction model iteration, plotting its true positive rate against the false positive rate across all classification thresholds; the convex, upward-bowed trajectory, with the curve consistently remaining above the diagonal baseline, indicates robust discriminative ability, suggesting an AUC in the approximate range of 0.70–0.75 and reflecting a meaningful capacity to distinguish between converted and non-converted accounts in the previous model.

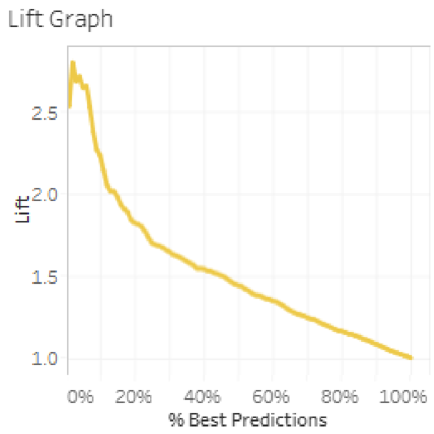

Figure 12: Previous Model: Lift Curve for Latest Version (100 Bins)

The lift curve for the latest upsell-prediction model, plotted across 100 equally sized prediction percentiles, exhibits a steep initial ascent, achieving a maximum lift of approximately 2.8 at the top 1% of ranked accounts, and then declines monotonically toward a lift of 1.0 at the 100th percentile; this profile indicates that the model concentrates conversion likelihood effectively in its highest-scoring deciles, with progressively diminishing enrichment as one moves down the ranking, reflecting robust early-decile discrimination in the previous model.

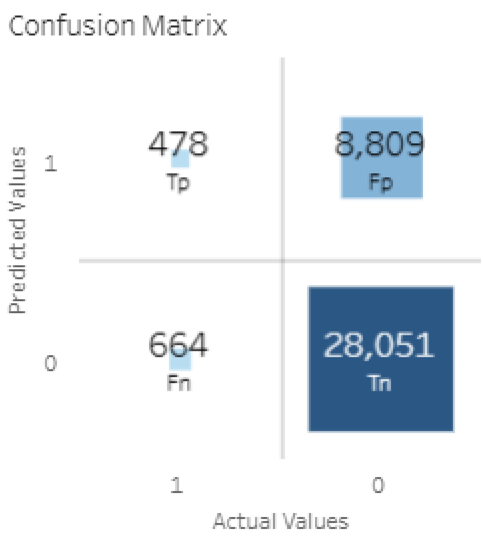

Figure 13: Confusion Matrix Analysis for Latest Version

The confusion matrix for the most recent upsell-prediction model shows 478 true positives and 28 051 true negatives alongside 8 809 false positives and 664 false negatives, indicating that while the model correctly identifies a substantial majority of non-converting accounts (TN rate ≈ 97.0%), it struggles with a relatively high false-positive rate (FP rate ≈ 23.9%) and a moderate false-negative rate (FN rate ≈ 58.1% of actual converters), reflecting a conservative threshold that prioritizes overall specificity over sensitivity in the previous model.