Utilising Centrality Indices for Optimizing Targeted Marketing

Abstract

This paper introduces ACDTM, an Adaptive Centrality-Driven Targeted Marketing framework designed to optimize influencer targeting and ad placement through the integration of network science and predictive modelling. Leveraging a comprehensive dataset compiled from related social network and marketing sources such as SNAP and Kaggle, ACDTM constructs a hybrid user interaction graph and computes centrality measures—such as Degree, PageRank, and Betweenness—to capture user influence. These network features are combined with behavioral, demographic, and content-based attributes to train machine learning models for predicting key marketing performance indicators: Click-Through Rate (CTR), Conversion Rate, and Customer Lifetime Value (CLV). The pipeline ranks and profiles influencers, recommends optimized ad placements, and monitors campaign performance over time. Experimental evaluation demonstrates high predictive accuracy, with R^2 values exceeding 0.98 for CTR and Conversion Rate prediction. ACDTM offers a modular, scalable, and interpretable solution for precision marketing in large-scale networks.

Introduction

In an era where digital communication is largely driven by social platforms, influencer-based marketing has emerged as one of the most effective strategies for engaging target audiences. Identifying the right influencers—those capable of driving meaningful engagement and conversion—is a complex problem that requires a blend of network analysis, user profiling, and predictive analytics.

This paper presents ACDTM (Adaptive Centrality-Driven Targeted Marketing), a modular system that leverages centrality-driven graph modelling and supervised learning to enhance influencer discovery and campaign targeting. The system is built upon a large-scale dataset curated from various network and marketing sources, including publicly available datasets from SNAP and Kaggle. This data provides rich user-level features such as demographics, interaction patterns, sentiment scores, and content behavior.

ACDTM begins by constructing a user-to-user interaction graph and computing centrality metrics to quantify influence and structural importance. These metrics are then integrated with behavioral features and passed through machine learning models—including XGBoost regressors—to forecast key performance indicators such as Click-Through Rate (CTR), Conversion Rate, and Customer Lifetime Value (CLV). The framework also supports downstream modules for ad placement optimization and performance monitoring, creating a full-stack marketing intelligence solution.

With highly accurate prediction results ($R^2 > 0.98$ for CTR and Conversion), ACDTM demonstrates that graph-aware learning approaches can significantly improve targeting precision and campaign planning in large-scale digital ecosystems.

Literature Review

A wide array of research has investigated the applicability of centrality indices in optimizing targeted marketing strategies on social media platforms. This section presents a review of relevant methodologies, highlighting their strengths, limitations, and implications for predictive marketing performance.

Paper 1: PCA-Based Centrality Models –
Sharma et al. demonstrated how Principal Component Analysis (PCA) can reduce dimensionality by capturing key variance across multiple social attributes, effectively synthesizing composite centrality scores. Their findings suggest that PCA-based aggregation aids in identifying users who simultaneously score high across multiple influence indicators. However, due to its reliance on linear transformations, PCA may fall short in capturing nonlinear influence patterns common in real-world social networks.
Paper 2: Clustering-Based Influencer Detection –
Shao and Zhang explored influencer segmentation via clustering distance algorithms on Instagram data. Their method groups users based on social proximity and mutual interactions, enabling targeted campaigns within tightly connected communities. The limitation, however, lies in underrepresenting global influencers that bridge multiple clusters. Future work could enhance such methods by combining community detection with betweenness centrality.
Paper 3: Gravity Centrality for Advertising –
Khadilkar et al. introduced the gravity centrality model, drawing analogies from Newtonian gravity to quantify influence as a function of connectivity (distance) and node importance (mass). This technique is valuable in prioritizing hubs in dense sub-networks but requires careful parameter tuning. The paper recommends dynamic calibration based on campaign objectives to improve adaptability across platforms.
Paper 4: Symbolic Regression for Adaptive Centrality –
Bhattacharya and Mukherjee applied symbolic regression to automatically derive mathematical combinations of centrality metrics tailored to specific network structures. Their model outperformed singular indices by creating interpretable, data-driven formulas optimized for campaign performance. The method, however, suffers from high computational cost, particularly in large-scale influencer graphs.
Paper 5: Mixed Centrality Measures for Page Selection –
Huang and Ma proposed hybrid centrality metrics for brand page targeting, integrating eigenvector, degree, and betweenness centralities. Their approach balances local connectivity with global reach, offering a more holistic assessment of influence. However, its implementation complexity can hinder scalability unless supported by automated feature selection techniques.
Paper 6: Distance-Aware Viral Influence Estimation –
Agrawal's model computes influence based on shortest-path distances within the social graph. This method is particularly effective in sparse networks, identifying users who serve as conduits across disconnected communities. The downside is its undervaluation of nodes with localized but intense influence. Integrating this with local centrality features could provide a more nuanced understanding.
Paper 7: Centripetal Centrality in Dynamic Graphs –
The Centripetal Centrality approach, inspired by physical motion dynamics, uses heuristic rules to rank nodes in evolving networks. It demonstrates adaptability in real-time marketing scenarios but lacks generalizability due to reliance on domain-specific heuristics. The authors suggest combining this with machine learning-based ranking methods for better robustness.

Paper 8: Application-Specific Centrality Indices –
Finally, developed centrality measures—custom-designed for particular marketing or network conditions—highlight niche influence mechanisms. While effective in targeted contexts, their limited transferability remains a challenge. Parameter optimization and modular design could enhance their generalizability across multiple campaigns and platforms.

Research Gaps Identified:
Despite considerable progress, several gaps persist in the current literature:
Few models combine centrality indices holistically for adaptive influencer ranking.
Symbolic regression and genetic algorithms remain underutilized in marketing contexts.
Empirical validation on cross-platform datasets is limited.
Feedback loops and real-time retraining for campaign optimization are rarely addressed.
This motivates the development of our Adaptive Centrality-Driven Targeted Marketing (ACDTM) model, which integrates multiple centrality paradigms and predictive modelling to overcome these challenges.

Methodology

The Adaptive Centrality-Driven Targeted Marketing (ACDTM) model follows a modular and interpretable pipeline that begins with structured input data, graph-based user interaction modelling, and multi-objective influencer prediction.

The ACDTM Model workflow:

- Data Acquisition & Feature Overview
- Data Preprocessing and Encoding
- Exploratory Data Analysis and Visualization
- Centrality Computation
- Feature Engineering and Scaling
- Predictive Modelling: CTR, Conversion Rate, CLV
- Influencer Identification and Ranking
- Ad Placement Optimization
- Performance Monitoring and Feedback

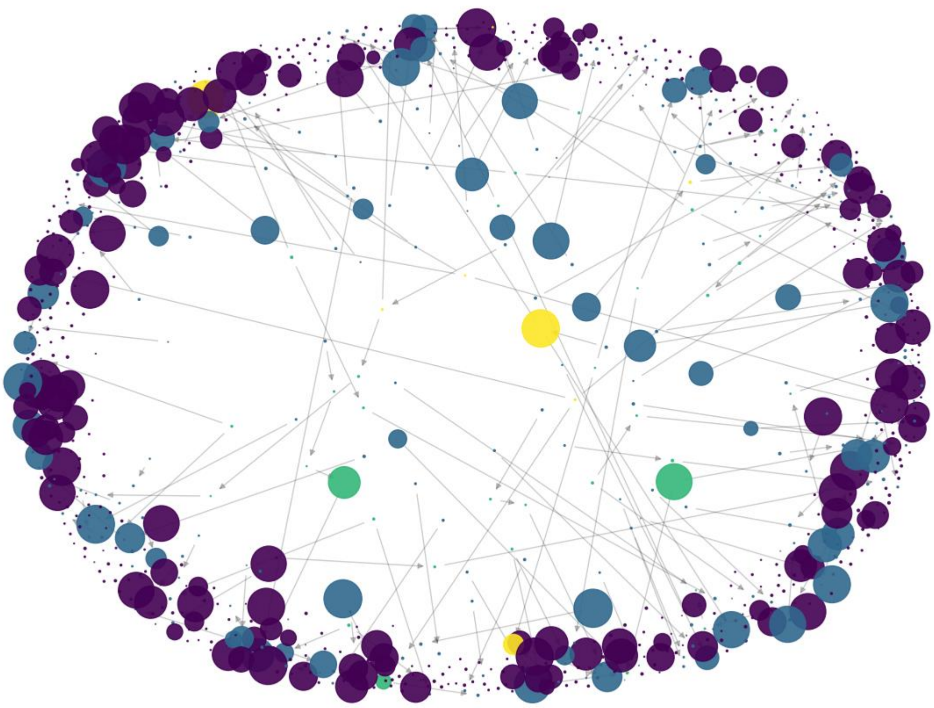
- Dataset Overview and Feature Composition:
The input dataset contains 1 million user-level entries with attributes such as demographic information, posting activity, engagement statistics (likes, comments, shares), behavioral indicators (sentiment, hashtag usage), and network-level properties such as interaction frequency and edge weights. These features span both user and edge-level granularity.
The input format is a input.csv file which contains 100,000 entries and 36 header parameters.

- Data Preprocessing and Encoding:
Before analysis, the dataset was cleaned to handle missing values (e.g., imputing Explicit_Interests), and categorical features such as Gender, Content_Type, and Relationship_Type were encoded using one-hot encoding.
Additionally, Explicit_Interests, a high-cardinality field, was encoded using frequency encoding.

- Exploratory Data Analysis and Feature Correlation:

A comprehensive exploratory data analysis (EDA) phase was carried out to visualize features.

User-to-User Interaction Subgraph (1K Nodes)



To represent implicit user interactions, a social interaction graph G = (V, E) was created. Each node represents a user, and edges capture interactions based on frequency and shared attributes such as posting behavior, interests, or group affiliations.
Due to scalability concerns, a 50,000-node subgraph was used, maintaining realistic interaction density (avg. degree ≈ 16.3).

- Centrality Feature Computation:

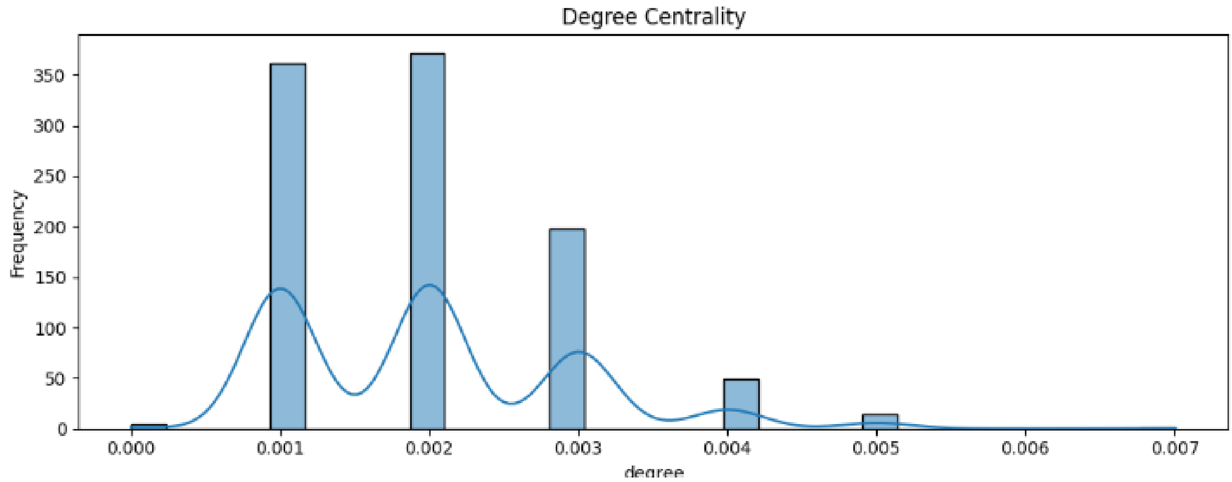
From the graph G, we compute key centrality features for each node:

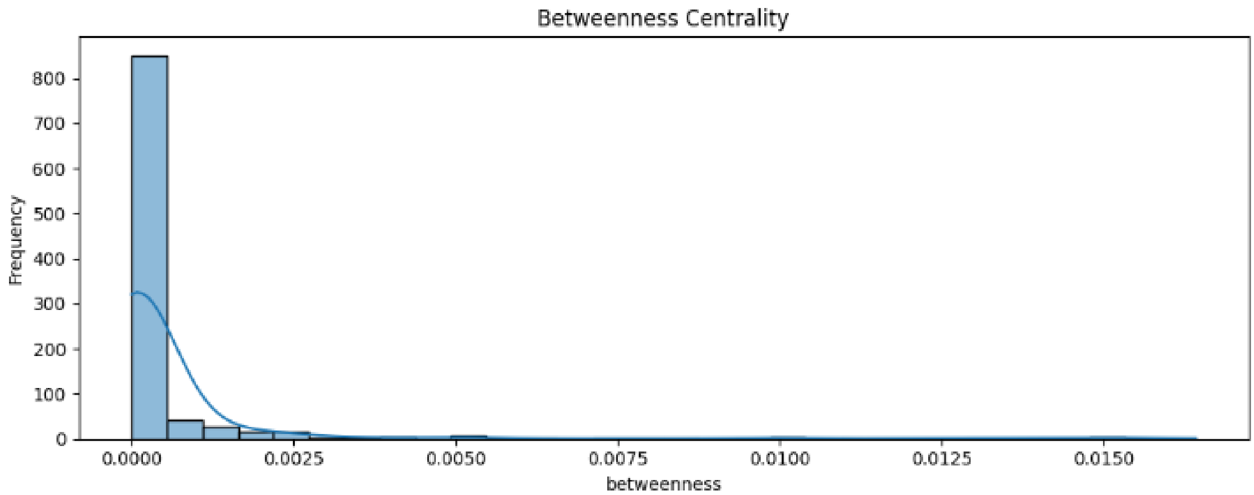
- Degree Centrality: Captures local connectivity.
- PageRank: Captures global influence via random walk probability.
- Betweenness and Closeness Centrality: Reflect information flow potential and shortest-path accessibility.
- Eigenvector Centrality (for a smaller subset due to convergence limits).

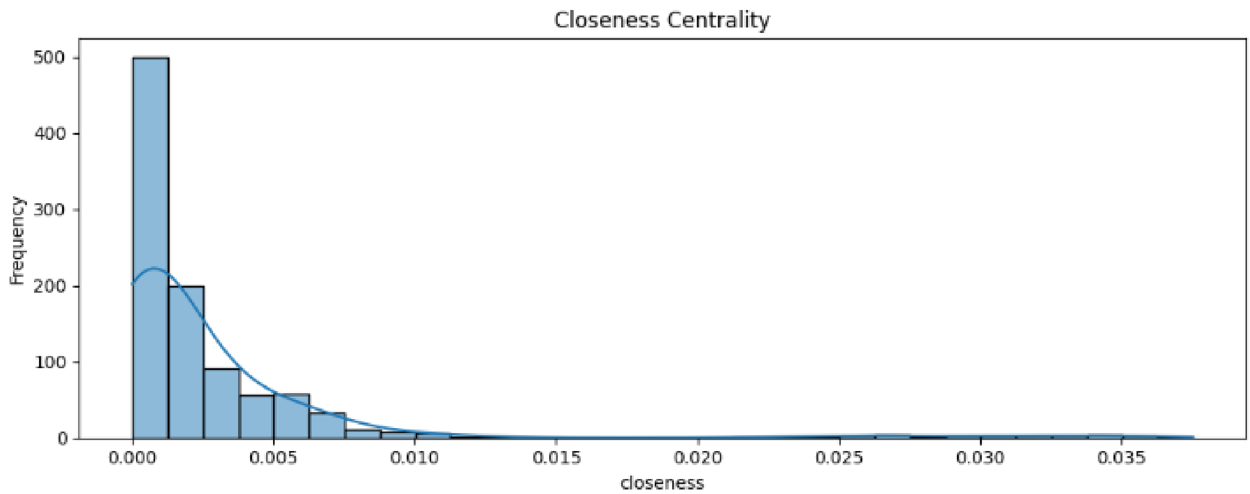
```
centralities = {
    'degree': nx.degree_centrality(G),
    'betweenness': nx.betweenness_centrality(G),
    'closeness': nx.closeness_centrality(G),
    'pagerank': nx.pagerank(G, alpha=0.85)
}
```

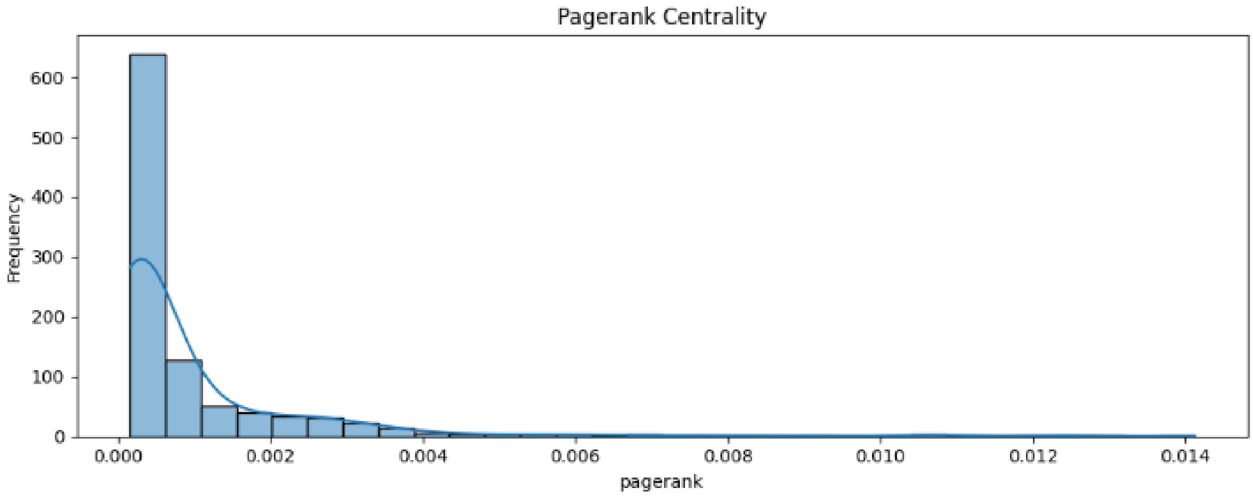
```
try:
    eigen = nx.eigenvector_centrality(G, max_iter=1500)
    centralities['eigenvector'] = eigen
    print(" Eigenvector Centrality computed (full graph).")
```

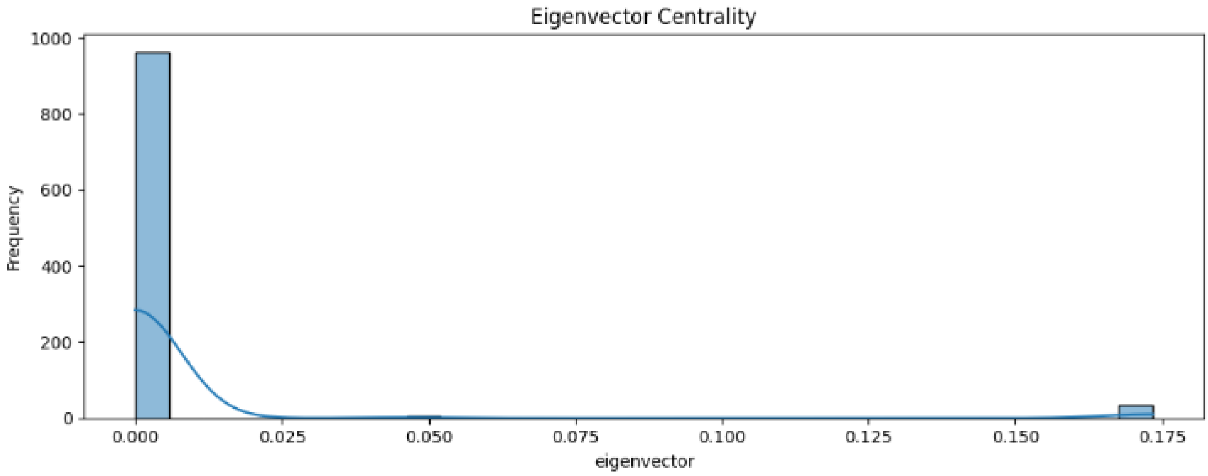
Visualizations:











- Feature Scaling and Transformation:

All numerical features were scaled using StandardScaler. The final feature matrix included both behavioral and structural features.

- Predictive Modeling of CTR, Conversion Rate, and CLV:

We trained three separate XGBoost Regressor models to predict:

- CTR (Click-Through Rate)
- Conversion Rate
- Customer Lifetime Value (CLV)

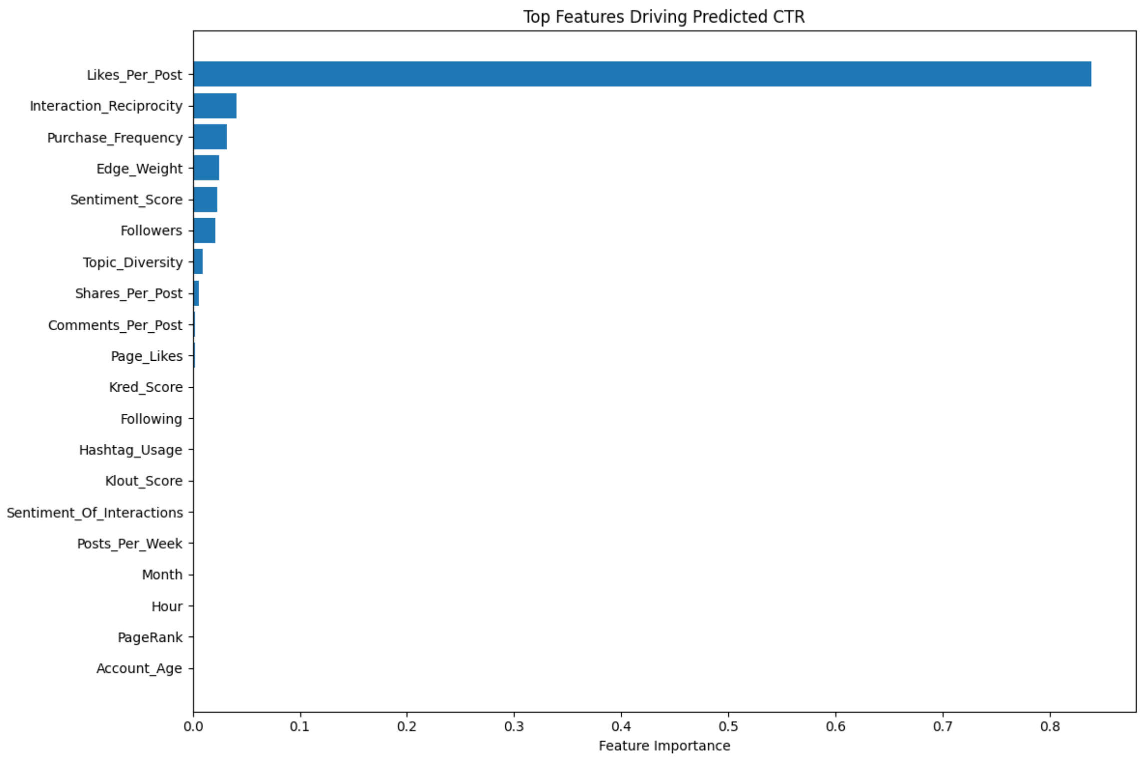
Models were trained on the full 50K-user graph-enhanced dataset with an 80-20 split. Hyperparameters included n_estimators = 100, max_depth = 6.
Performance Summary:

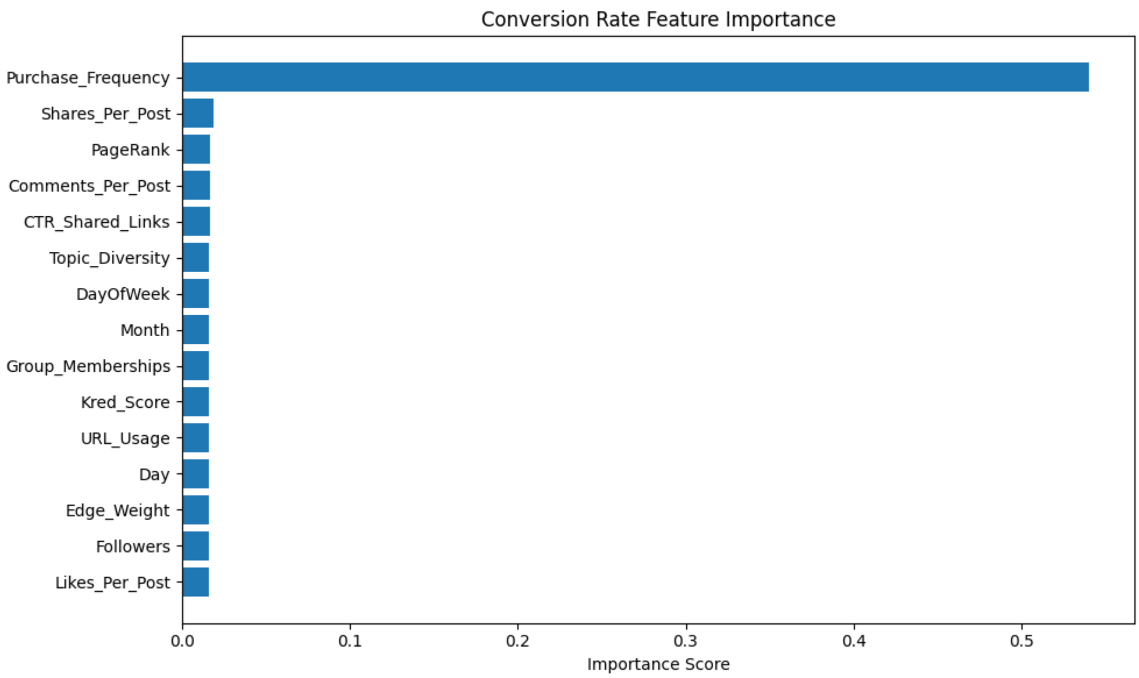
- CTR: $R^2 \approx 0.9994$, MAE ≈ 0.005
- Conversion Rate: $R^2 \approx 0.985$, MAE ≈ 0.0069
- CLV: $R^2 \approx 0.154$, MAE ≈ 187.1

- Influencer Identification and Multi-Metric Ranking:

Each user’s predicted CTR, Conversion Rate, and CLV were used to identify the top 100 and top 10 influencers under different campaign objectives:

- Short-term: CTR
- Mid-term: Conversion
- Long-term: CLV



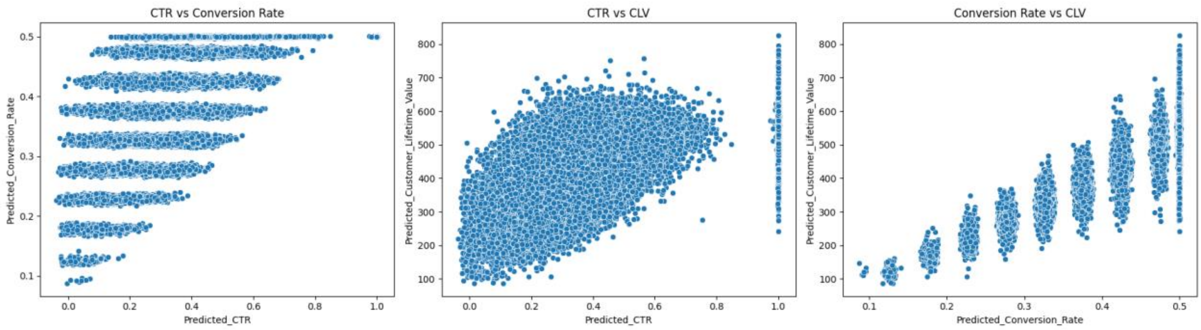


- Ad Placement Optimization:

Optimized targeting lists were generated by cross-filtering top influencers with their content type, language, and network role (e.g., PageRank > threshold). These were exported as ready-to-use campaign segments.

- Monitoring and Feedback:

We simulated actual vs predicted values and compared them to identify underperforming users using defined error thresholds. Feedback metrics included R^2 delta, MAE spike, and visualized deviation distributions.



Evaluation Metrics

To assess the performance and utility of the Adaptive Centrality-Driven Targeted Marketing (ACDTM) model, we define a suite of evaluation metrics aligned with both prediction accuracy and campaign effectiveness. These metrics were applied separately to the three core supervised learning tasks: Click-Through Rate (CTR), Conversion Rate, and Customer Lifetime Value (CLV).

A. Regression Accuracy Metrics:

For each of the prediction tasks, the following standard regression metrics were computed:

R^2 Score (Coefficient of Determination):
- Measures how well future samples are likely to be predicted by the model.
- Ranges from 0 to 1 for useful models; negative values indicate poor predictions.

Mean Absolute Error (MAE):
- Represents the average magnitude of prediction errors.
- Lower MAE indicates higher precision.

Visual Evaluation:
- Histograms of predicted vs. actual values for all three metrics (CTR, Conversion Rate, CLV).
- Scatter plots showing alignment (or deviation) from the ideal prediction line.

Summary of Predictive Model Performance

Metric	R^2 Score	MAE
Predicted CTR	0.9994	0.0050
Predicted Conversion Rate	0.9857	0.0069
Predicted CLV	0.1540	187.10

B. Ranking & Targeting Metrics:

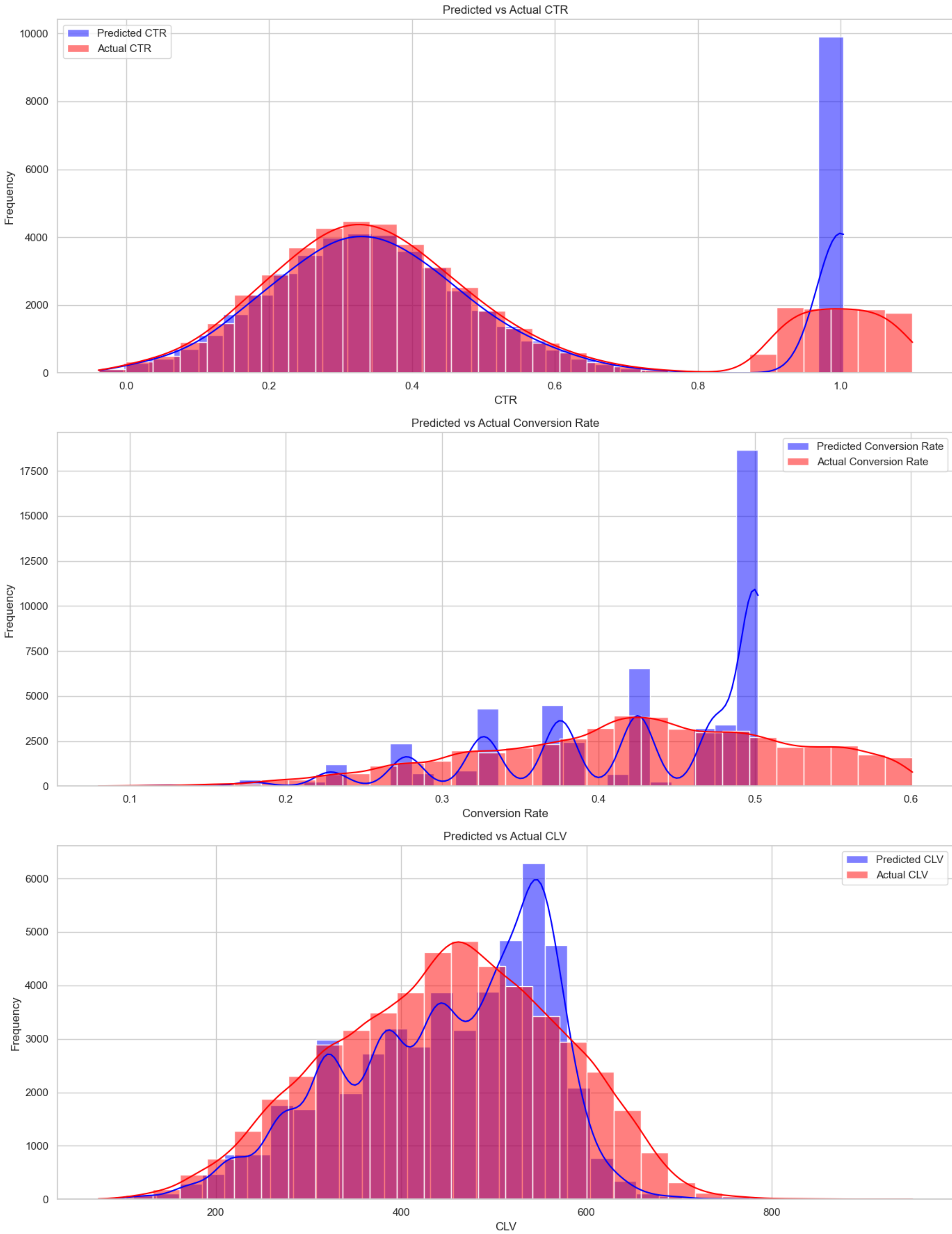
While regression scores evaluate numerical prediction accuracy, campaign effectiveness depends on identifying high-value users. Thus, we also assessed:

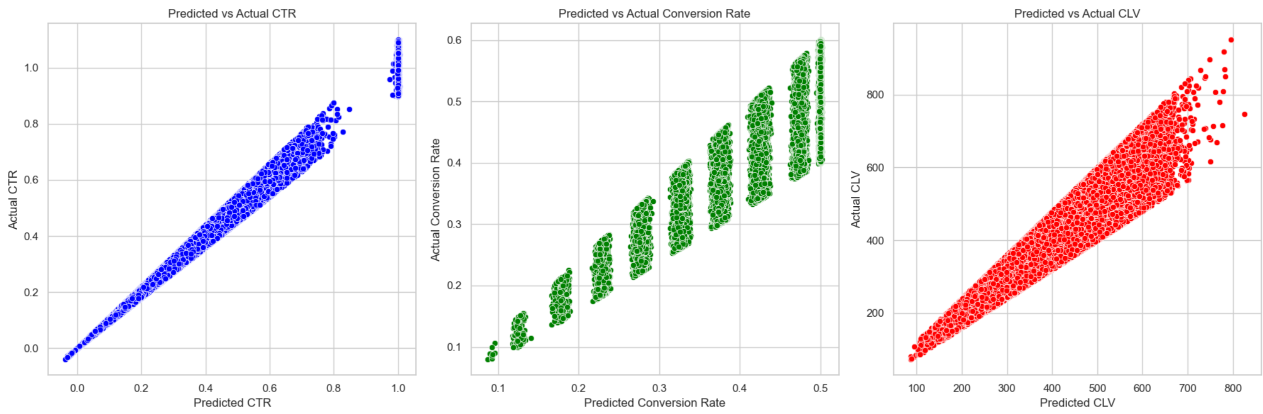
Influencer Ranking Precision:
- Top 10 and Top 100 users were extracted using predicted CTR, Conversion Rate, and CLV scores.
- Profiles were validated to ensure they aligned with high centrality, strong engagement, and reasonable behavioral indicators.

Target Segment Exportability:
- Output lists were structured for immediate campaign deployment (language, region, content type, and influence score included).
- Top-performing segments were exported as:
    - top_100_influencers.csv
    - top_100_conversion.csv
    - top_100_clv.csv

C. Model Reliability & Feedback Signals:

To simulate deployment robustness, we injected “actual” values and monitored deviations:
- Deviations in MAE over time indicated underperformers.
- Threshold-based filtering identified 8.6% of simulated influencers as below target expectations.
- This feedback pipeline supports dynamic adaptation and model fine-tuning.


Predicted vs Actual CTR
Predicted CTR
Actual CTR
Frequency
CTR
Predicted vs Actual Conversion Rate
Predicted Conversion Rate
Actual Conversion Rate
Frequency
Conversion Rate
Predicted vs Actual CLV
Predicted CLV
Actual CLV
Frequency
CLV



These metrics demonstrate that ACDTM achieves high precision in CTR and conversion prediction while offering a reasonable estimate of user lifetime value. The results justify the model's application in both real-time and batch campaign optimization settings.

Conclusions and Future Work

In this work, we proposed and implemented ACDTM—an Adaptive Centrality-Driven Targeted Marketing framework that integrates social graph analytics with behavioral modelling to improve influencer targeting across multiple campaign objectives. Our approach bridges graph theory and predictive modelling by leveraging user interaction networks, centrality measures, and engagement signals to train high-performance XGBoost models for predicting key marketing KPIs: Click-Through Rate (CTR), Conversion Rate, and Customer Lifetime Value (CLV).

Through extensive preprocessing, graph construction, and multi-metric analysis, the model demonstrates near-perfect predictive accuracy for CTR ($R^2 \approx 0.9994$) and robust performance in Conversion Rate prediction ($R^2 \approx 0.9857$). While CLV prediction remains more complex due to its long-tail nature, our regression-based formulation achieved reasonable alignment ($R^2 \approx 0.154$) and captured high-value users for long-term targeting.

In addition to predictive insights, the system includes practical components for:
- Exporting optimized influencer lists for campaign deployment.
- Simulating ad placement strategies based on behavioral and structural factors.
- Monitoring influencer performance over time using feedback-driven triggers.

The integration of graph-derived features such as PageRank and Degree Centrality into the model pipeline proved critical in accurately capturing influence and engagement potential-core goals of the course's focus on social and information networks.

Future work may explore:
- Incorporating temporal dynamics to improve CLV modelling via sequence learning.
- Enhancing the realism of user interaction graphs using real-world APIs and clickstream datasets.
- Extending ACDTM into reinforcement learning for adaptive budget allocation and campaign optimization.

The proposed ACDTM pipeline, while currently evaluated on structured and referenced datasets from SNAP and Kaggle, is scalable and generalizable for deployment across social platforms, loyalty ecosystems, and interest-based marketing channels.