***TLADS Step 6: Explore Analytic Algorithms***

|  |  |  |  |
| --- | --- | --- | --- |
| **Business Initiative:**  Improve user retention and personalised engagement strategies for Centralised Crypto Exchanges (CEXs - user retention, churn prediction and behavioural segmentation) to increase fee revenue. | | | |
| **Prioritised Use Case:**  User segmentation based on wallet behavioural archetypes. | | | |
| **Analytic Score** | **Score Explanation** | **Analytics Algorithm** | **Algorithm rational** |
| Innovator Propensity Score | Measures likelihood of exploring new protocols, contracts or novel products. | K-Means clustering (unsupervised) | Identifies wallets that explore novel clusters of methods / protocols (interpretability: moderate, good for unsupervised) |
| Random Forest Classifier | Offers feature importance and handles heterogenous data well (interpretability: high) |
| Autoencoder-based anomaly score | Flag outliers as innovators if their behaviour deviates significantly from the norm – ideal if your using embeddings. |
| Revenue Contribution Score | Estimated economic contribution based on usage and transfer metrics | Gradient Boosting Regressor | Balances accuracy and performance; good for noisy financial data. |
| Linear Regression | Gives simple, explainable drivers of revenue (good stakeholder transparency). |
| Clustering + Aggregated Volume Bins | Can reveal tiers (whales, minnows, etc.) for flat fee personalisation. |
| Engagement Stability Score | Quantifies how reliably active a user is over time | Hidden Markov Models | Directly models latent user states – useful for loyalty design. |
| Rolling Window Statistics + Decision Trees | Allow quick thresholds like “90 days active” – easy business logic. |
| Time-Series K-Means Clustering | Captures seasonal vs consistent vs episodic usage. |
| Product Affinity Score | Measures alignment to categories like DeFi, NFTs, tokens | Multi-class Logistic Regression | Keeps the score linear + interpretable. |
| PCA + K-Means on event category proportions | Handles dense feature spaces. |
| Topic Modelling (LDA) on interaction sequences | Works well for unordered wallet interaction “documents” (each method is a “word”) |
| UX Complexity Tolerance Score | Measures comfort with complex or multi-step actions. | Decision Tree with protocol / method diversity thresholds | Ideal for threshold-based UX routing. |
| Neural Net Regression (for score from many sparse inputs) | Can model non-linear relationships for UX segments. |
| Feature Entropy Calculation (as proxy for cognitive load) | Is easily computed, explainable and maps to UX theory. |
| Interaction Mode Score | Dominant modes of interaction (e.g., bridges vs DEXs) | Hierarchical Clustering on normalised event factors | Can capture nuanced behaviour groupings. |
| Spectral Clustering | Can capture nuanced behaviour groupings. |
| t-SNE + Manual Labelling for UX personas | For visual prototyping — key to UX tuning via stakeholder input |
| Reward Responsiveness Score | Predicts change in activity after past incentives (or modelled responsiveness). | Uplift Modelling (Two-model approach) | directly measures treatment/control lift. |
| Casual Forests | (meta-learners) are more rigorous for incentive impact |
| Time Series Regression with exogenous variables | captures lagged reward effect on activity. |
| Behavioural Stickiness Score | Measures inertia in dApp/method usage. | Normalised Entropy | Simple, transparent. |
| Cosine Similarity vs Historical Vectors | shows change from historical usage vectors. |
| Lasso Regression on event history | shrinks less-used features → shows key behavioural anchors. |
| LTV Proprietary Score | Predicted cumulative value from a wallet over its lifetime. | Gradient Boosting Regression | accurate and robust to outliers. |
| Survival Analysis | Models help model “time to churn” alongside value. |
| Multivariate Time-Series Models (VAR, Prophet) | Time-series adds seasonality and future projection ability. |
| Dormancy Risk Score | Likelihood that a user will become or remain inactive. | Logistic Regression | Fast, explainable |
| Time-Series Anomaly Detection (Rolling Z-Score) | Z-score is simple and actionable |
| LSTM Binary Classifier (if sequence data is rich) | LSTM enables nuanced sequential detection of drop-off. |
| Bridge Utilisation Score | How actively the user uses cross-chain bridges. | MinMax Scaled Volume Model | Simple scoring helps CEX prioritise by bridge size. |
| K-Means or DBSCAN on bridge counts and volume | DBSCAN helps detect bridge-heavy clusters regardless of distribution |
| Quantile Scoring |  |
| Cross-Domain Engagement Score | Measures breadth across DeFi, NFTs, tokens, etc. | Diversity Index (e.g., Shannon Entropy) | Entropy is simple and expressive. |
| SoftMax-normalised scores from engagement vector. | SoftMax gives probabilistic “domain affinity” distribution. |
| PCA + K-Means | PCA can reduce noise across high dim inputs. |
| Segment Membership Score | Classifies wallet into defined user personas. | Multiclass Classifier (XGBoost, RF, Logistic Regression) | Multiclass classifier = direct segment prediction. |
| Latent Dirichlet Allocation (for behaviour sequence modelling) | **LDA** gives soft segment distributions (user overlaps). |
| Graph Clustering (wallet similarity graph) | Graph methods capture behaviour similarity, e.g. co-bridge usage. |
| Behavioural Volatility Score | Measures changeability in behaviour over time. | Rolling Standard Deviation of Event Categories | Rolling std = fast signal of volatility. |
| Time-Series Clustering | **Clustering** helps surface seasonal vs erratic users. |
| Variance Feature Importance | Useful to adjust message frequency and ad cadence. |