PeakPicks

PeakPicks.ai

Leveraging Advanced AI for Profitable Product Discovery on Amazon

Executive Summary

In the dynamic and highly competitive marketplace of Amazon, identifying profitable and promising products is a significant challenge for investors and sellers alike. PeakPicks.ai addresses this challenge by employing advanced artificial intelligence models to forecast product performance accurately. Our classification model achieves a quality of 90%, our regression model attains a Weighted Absolute Percentage Error (WAPE) of 15–20% (depending on the product category), and our sales estimator achieves a quality of 97%. This white paper delves into the technical intricacies of our two-step forecasting process, machine learning frameworks, and the innovative methodologies that set us apart in the industry.

Our approach combines classification and regression analyses to not only predict which products are likely to experience sales growth but also quantify the extent of that growth. By utilizing gradient-boosted trees, optimizing for appropriate loss functions like Tweedie loss, and employing metrics such as WAPE and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC), we enhance prediction accuracy significantly.

Additionally, we introduce a sales estimator that maps predicted sales ranks to actual sales figures using a specialized regression model. This comprehensive framework enables us to provide actionable insights for investment decisions, ultimately maximizing returns and understanding market dynamics more profoundly.

Table of Contents

Introduction

In an era where e-commerce platforms like Amazon host millions of products, identifying which items are likely to yield profitable returns is akin to finding a needle in a haystack. Traditional methods of product selection often fall short due to the sheer volume of data and the complexity of market dynamics. PeakPicks.ai emerges as a solution, leveraging advanced artificial intelligence (AI) and machine learning (ML) techniques to provide precise and actionable product recommendations.

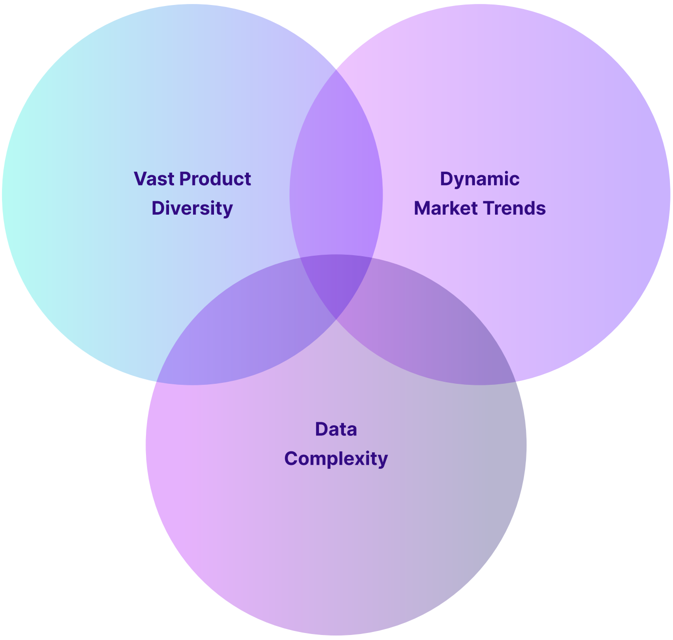
This white paper explores the technical foundation of PeakPicks.ai's forecasting models, detailing the methodologies, algorithms, and metrics that drive our recommendations. By dissecting our two-step forecasting process and the machine learning tasks involved, we aim to provide a comprehensive understanding of how we harness data to predict product performance on Amazon.

The Challenge

The Amazon marketplace is characterized by:

- Vast Product Diversity: Millions of products across countless categories.
- Dynamic Market Trends: Rapid changes in consumer preferences and competitive actions.
- Data Complexity: An overwhelming amount of unstructured and structured data.

Figure 1: Challenges in the Amazon Marketplace



These factors make it challenging for investors and sellers to:

- Identify products with high growth potential.
- Forecast future sales accurately.
- Make informed investment decisions to maximize returns.

Traditional analytical methods are insufficient for handling the scale and complexity of the data, necessitating advanced AI-driven solutions.

Our Objective

Our primary goal is to identify products worth investing in by predicting their future sales performance. This involves:

- Determining Potential Winners: Classifying products that are likely to experience an increase in sales.
- Quantifying Growth: Estimating the magnitude of the expected sales increase for each product.

By achieving these objectives, we enable investors and sellers to focus on products with genuine growth potential, thereby maximizing returns and optimizing investment strategies.

The Two-Step Forecasting Process

Figure 2: Overview of the Two-Step Forecasting Process

Our forecasting model is structured into two interconnected steps.

Classification Step

Objective:
Determine whether a specific product (ASIN) will experience an increase in sales compared to its current performance over a future period of k days.

Approach:

- Utilize a classification model to assign a growth probability to each product.
- Focus on identifying products with the highest likelihood of sales improvement.

Performance:
Our classification model achieves a quality of 90%, indicating high reliability in identifying potential growth products.

Regression Step

Objective:
Quantify the expected increase in sales for the products identified in the classification step.

Approach:

- Employ a regression model to estimate the exact amount by which a product's sales will increase.
- Provide a numerical forecast of potential growth levels for each product.

Performance:
The regression model attains a Weighted Absolute Percentage Error (WAPE) of 15–20%, depending on the product category, demonstrating strong predictive accuracy.

Rationale:
Combining both steps enhances the accuracy and reliability of our predictions. Relying solely on regression would lack the initial filtering of potential winners, while classification alone would not provide the extent of growth necessary for informed decision-making.

Machine Learning Framework

Our machine learning tasks align with the two steps:

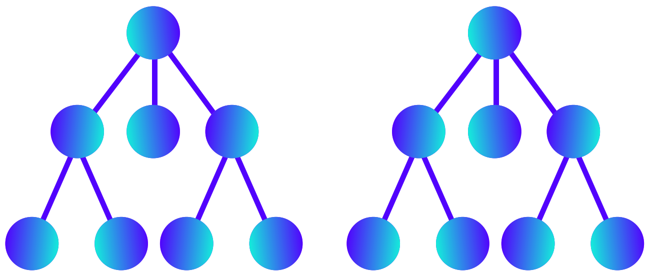
- **Classification Task:** Predict a categorical outcome—whether a product's sales will increase.
- **Regression Task:** Predict a continuous outcome—the magnitude of the sales increase.

Gradient-Boosted Trees

We utilize gradient-boosted trees as our primary modeling framework due to their:

- **Handling of Complex Datasets:** Effective with non-linear relationships and interactions.
- **Robustness:** Resistant to overfitting with proper regularization.
- **Flexibility:** Suitable for both regression and classification tasks.

Figure 3: Architecture of Gradient-Boosted Trees

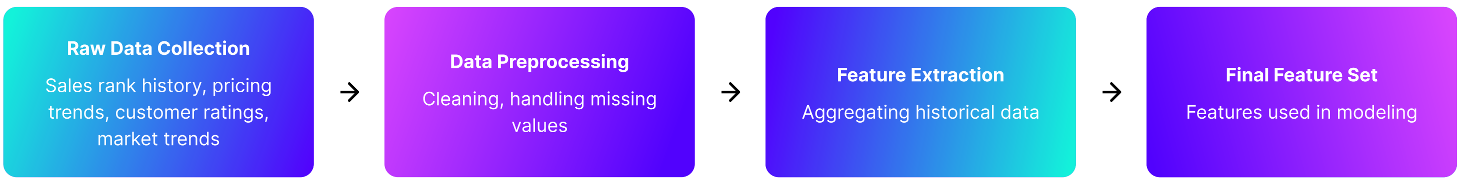


Feature Engineering

We aggregate features derived from historical data over specified time periods, including:

- **Sales Rank History**: Trends and fluctuations in product rankings.
- **Pricing Trends:** Historical pricing data and discount patterns.
- **Customer Ratings:** Reviews and rating trajectories.
- **Market Trends:** Category-specific dynamics and seasonal effects.

Figure 4: Feature Engineering Pipeline



These features capture the multifaceted factors influencing product performance.

Regression Analysis

Tweedie Loss Function

During regression model training, we optimize directly for the Tweedie loss function because:

- **Heavy-Tailed Distributions:** Sales data often exhibit heavy tails, with extreme values that can skew results.
- **Flexibility:** Tweedie loss accommodates distributions that are between Poisson and gamma, suitable for modeling sales data.

Formula:
The Tweedie loss function is defined as:

$$L(y,\, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\hat{y}_i^{1-p}}{1-p} - \frac{y_i \hat{y}_i^{-p}}{-p} \right)$$

where:

- y_i is the actual value.
- $\hat{y}_i$ is the predicted value.
- p is the Tweedie parameter (1 < p < 2 for compound Poisson-gamma distribution).

Weighted Absolute Percentage Error (WAPE)

Our primary metric for regression evaluation is WAPE, calculated as:

$$\text{WAPE } = \frac{\sum_{t=1}^{n} |F_t - A_t|}{\sum_{t=1}^{n} A_t} \times 100\%$$

where:

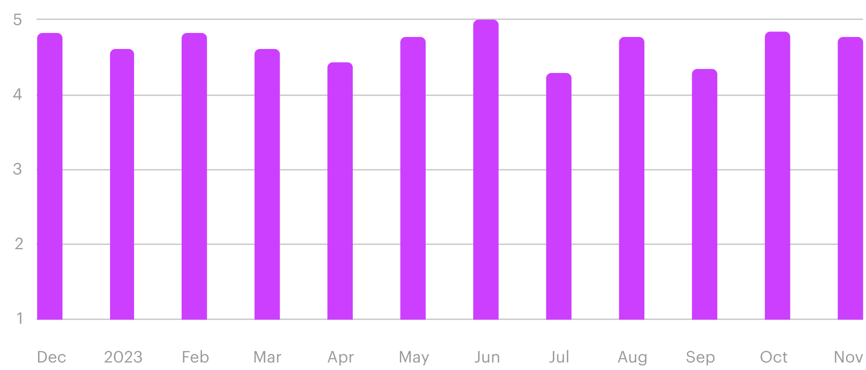
- F_t is the forecasted value at time t.
- A_t is the actual value at time t.

Performance:
The regression model achieves a WAPE of 15–20%, varying by category. This low percentage indicates strong predictive accuracy in estimating sales increases.

Comparison with Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\%$$

Figure 5: Comparison of WAPE and MAPE Metrics



Limitations of MAPE:

- **Penalizes Small Targets:** Errors when A_t is small inflate the MAPE disproportionately.
- **Division by Zero**: Undefined when actual values are zero.

Why WAPE is Preferred:

- **Weighted Approach:** Aggregates absolute errors weighted by actual sales, mitigating the impact of small A_t values.
- **Stability:** Provides a more reliable measure for skewed data distributions.

Classification Analysis

Logloss Optimization

During classification model training, we optimize for the logarithmic loss (logloss) function:

$$\text{Logloss} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right]$$

where:

- y_i is the actual class label (0 or 1).
- p_i is the predicted probability of the positive class.

Advantages:

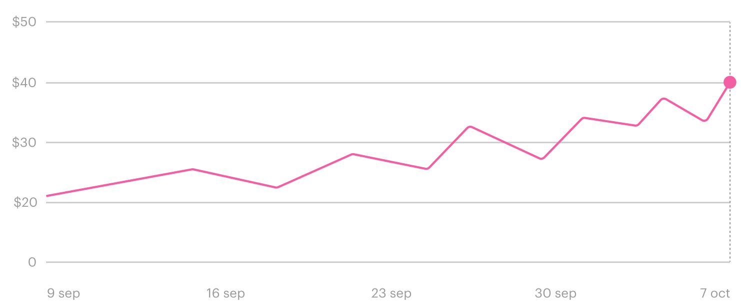
- **Probabilistic Interpretation:** Encourages models to output calibrated probabilities.
- **Sensitivity:** Penalizes confident but wrong predictions more severely.

ROC-AUC Metric

Our evaluation metric for classification is the Receiver Operating Characteristic–Area Under the Curve (ROC-AUC).

Performance: The classification model achieves a quality of 90%, reflecting its high reliability in distinguishing between products that will or will not experience sales growth.

Figure 6: ROC Curve Illustrating Model Performance



Advantages:

- **Threshold-Independent:** Evaluates model performance across all classification thresholds.
- **Interpretability:** A higher AUC indicates better model discrimination between classes.

Combining Both Analyses

Intersection of Top ASINs

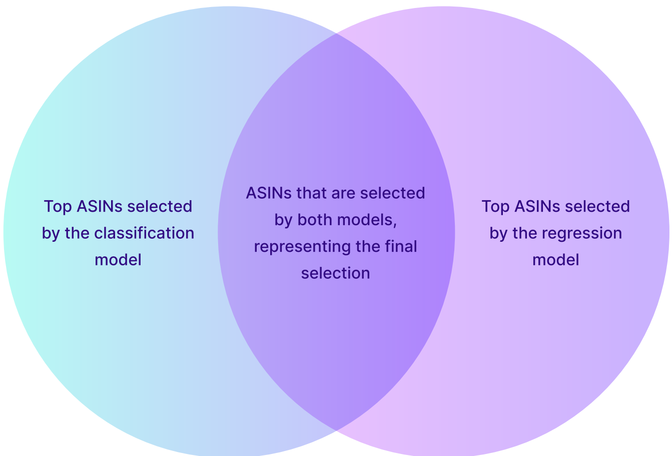
To synthesize the insights from both models:

1. **Select Top ASINs from Classification:** Products with the highest growth probabilities.
2. **Select Top ASINs from Regression**: Products with the highest predicted sales growth.

Post-Processing Techniques

- **Intersection:** Identify ASINs that appear in both top lists, ensuring they are both likely to grow and expected to grow significantly.
- **Ranking:** Prioritize ASINs based on combined scores from both models.
- **Filtering:** Apply business rules or constraints (e.g., inventory levels, category preferences).

Figure 7: Venn Diagram of ASIN Selection



Outcome:

- A robust list of ASINs with the greatest potential for success.
- Enhanced decision-making by focusing on products that meet both criteria.

Sales Estimator

Mapping Sales Ranks to Sales

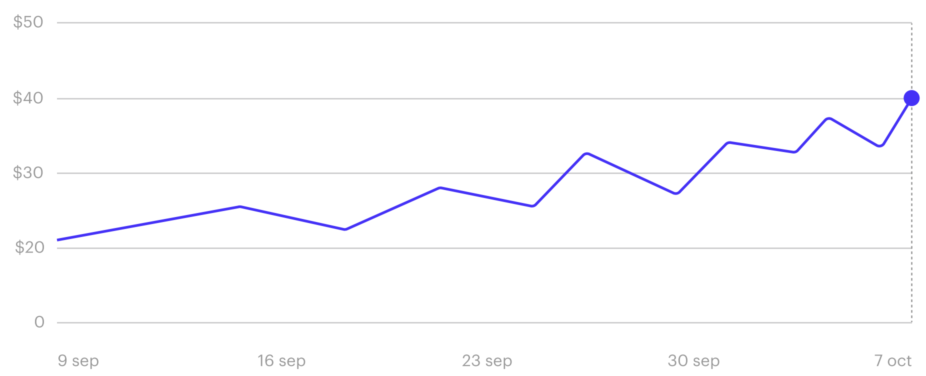
We develop a sales estimator to convert predicted sales ranks into actual sales figures.

Approach:

- **Data Collection:** Gather pairs of sales ranks and corresponding sales within each category.
- **Regression Modeling:** Use a lightweight regression model to predict sales based on sales rank.

Performance:
The sales estimator achieves a quality of 97%, effectively mapping sales ranks to actual sales with high precision.

Figure 8: Sales Rank vs. Sales Regression Curve



Regression Model for Sales Estimation

- **Interpolation:** Allows for estimating sales for ranks not explicitly observed in the data.
- **Smoothing Outliers:** Reduces the impact of anomalous data points that could skew average calculations.
- **Model Selection:** Choose models that capture the relationship between rank and sales accurately (e.g., logarithmic or exponential models).

Advantages over Simple Aggregation

- **Precision:** Provides more accurate sales estimates than category-wide averages.
- **Flexibility:** Adapts to changes in the relationship between rank and sales over time.

Benefits and Advantages

Enhanced Accuracy

The two-step forecasting process improves prediction precision, with our classification model achieving 90% quality and our regression model attaining a WAPE of 15–20%.

Informed Decision-Making

Quantitative estimates of sales growth, backed by high-quality models, aid investment strategies.

Market Insight

Understanding product trajectories helps navigate market dynamics.

Resource Optimization

Focusing on high-potential products maximizes return on investment, supported by our sales estimator's 97% quality.

Adaptability

The framework can be updated with new data, keeping predictions current and maintaining high model performance.

Conclusion

PeakPicks.ai offers a sophisticated and effective solution for identifying profitable and promising products on Amazon. By integrating classification and regression models that achieve high performance metrics (90% classification quality, 15–20% regression WAPE, and 97% sales estimator quality), optimizing for appropriate loss functions, and employing robust evaluation metrics, we provide accurate forecasts of product performance.

Our innovative methodologies, including the sales estimator and the intersection of model outputs, set us apart in the marketplace. We empower investors and sellers with actionable insights, enabling them to make informed decisions and capitalize on market opportunities.

As the e-commerce landscape continues to evolve, PeakPicks.ai remains committed to refining our models and expanding our capabilities, ensuring we stay at the forefront of AI-driven product recommendation technology.

Contact Information

For more information about our technology and services, please visit www.peakpicks.ai or contact us at contact@peakpicks.ai.