Smart Predictor for Smartphone Prices

Jernai Bennett
Florida International University
11200 SW 8th St
Miami, FL 33199
17865930501
Jbenn058@fiu.edu

Luis Fabrizio Barrera Inga
Florida International University
11200 SW 8th St
Miami, FL 33199
19542534953
lbarr171@fiu.edu

ABSTRACT
Choosing which smartphone to buy can be difficult for buyers that lack knowledge of the smartphone market. Some buyers may accidentally overspend on a smartphone with features that they don't need. Since phones can be expensive, it is especially important for buyers to make sure that they are buying a smartphone with all the features that they want at a decent price. Our model can help buyers accomplish this. Using smartphone data from two datasets, we trained our model to help buyers better understand how certain features inflate the price of smartphones.

CCS Concepts
- **Computing methodologies → Machine learning → Machine learning approaches → Supervised learning → Regression**
- **Information systems → Information systems applications → Data mining → Predictive analytics.**

Keywords
smartphone, dataset, feature, model, market

1. INTRODUCTION
Our project, "Mobile" Match, is a smart predictor for smartphone prices. Our project's goal is to solve the problem buyers face when choosing from hundreds of smartphone options. With almost 1000 models in the real world, from dozens of brands such as Apple and Tesla, finding the correct phone with the right features at the right price can be overwhelming for certain buyers.

1.1 Problem Definition
Almost every person in the world has a smartphone. Smartphones allow people to do everyday tasks like messaging family members and browsing social media. Buying a smartphone is expensive, so choosing the perfect phone can be challenging for buyers, especially those who do not have that much technical expertise. A buyer will want to buy a smartphone with all the features they want at a decent price. But sometimes, certain buyers can make the mistake of overspending on certain features that they don't need, for example, an advanced camera. That's why a customer needs to have knowledge of the features and prices that come with certain smartphone models.

Most people usually just buy from bigger companies like Apple and Samsung, without even considering other companies that could have better deals. This is where our model comes into play. Our model predicts smartphone prices and recommends devices that match both technical needs and budget constraints. By analyzing all the smartphones in our database, Mobile Match helps consumers make confident purchases without requiring technical expertise. Our model does this by translating complex specifications into everyday language that most people can understand. This helps users understand the practical benefits of every smartphone option and avoid overspending on features that they do not need.

Our model's goal is to simplify the smartphone selection process by analyzing specification-price relationships across markets, comparing regional pricing variations across markets, developing predictive models that can estimate pricing based on specifications in different regions, identifying key value drivers in smartphone pricing, and creating a data-driven framework for comparing smartphones across different brands and price segments. By completing these objectives, our model can help eliminate the confusion caused by the vast array of smartphone options and provide a clearer view of the global smartphone market.

DOI: http://dx.doi.org/10.1145/12345.67890

2. DATASET DESCRIPTION
The first dataset that we used to train our model is called" Mobiles Dataset 2025.csv." We found this dataset on Kaggle. This dataset had a total of 930 different smartphone models from companies all over the world such as Apple and Oppo, which is a Chinese based electronics company.

The dataset had a total of 15 columns/attributes that described basic information for every smartphone model. Some of the attributes that stood out to us were "RAM," "Launched Year," and five different launch price columns for five different markets. These markets are Pakistan, India, China, USA, and Dubai. The "RAM" column was key for us since storage is obviously an especially important characteristic that buyers care about. Buyers are most likely interested in phones with lots of storage space. The "Launched Year" column was key to us since it allowed us to see how many of the phones were newer and how many were a bit older. The five different launched price columns allowed us to compare smartphone prices across each market. We also wanted to discover which smartphone attributes were more impactful on smartphone prices in each market.

2.1 Preprocessing
After finding the dataset, we now needed to implement data preprocessing. We first decided to deal with null values. To our surprise, this dataset did not have a single null value in any of its 930 different smartphone models, which is a huge positive. Our

next task was to remove the columns that we felt were not as important as the other columns. Two columns that we removed were the front camera column and mobile weight column. We felt that when customers go to buy a smartphone, they don't really care about these features. For example, it is rare that a customer will specify that they want to buy a smartphone that weighs less than 6.5 ounces. Out of the five different markets, we decided to remove the prices for Pakistan and Dubai.

We wanted to focus on the three bigger markets, which were USA, China, and India. After dimensionality reduction, we ended up with these columns: "Company Name," "Model Name"," RAM," "Back Camera," "Processor," "Battery Capacity," "Screen Size," "Launched Price (India)," "Launched Price (China)," "Launched Price (USA)," and "Launched Year." At first, we removed the processor column, but then we realized that the processor type could have some importance in affecting the price. A high-quality processor costs more than a low-quality processor.

We wanted to train our model on smartphones that are easy to find and available in all three markets. If a smartphone were created around five years ago, it may not be sold anymore and if it is, it could be extremely hard to buy. Because of this, we removed any smartphone that was made before 2020. We also removed some of the smaller smartphone brands since these brands are not available in all of three major markets. We kept the brands that sell smartphones in the USA, China, and India. Now our data has phones that are more likely to be found in stores in the USA, China, and India.

The next step was to clean up all the values in the cells. Unfortunately, all the cells in this dataset had units, which meant that these cells had string values. For example, a smartphone could have "6GB" in the RAM column and "3,600mAh" in the battery capacity column. Machine learning models can only train on numeric values and not string values. Some cells even had two values. For example, a smartphone could have "50MP + 12MP" in its back camera column. Therefore, we used five functions: one that cleaned up the battery capacity column, another one that cleaned up the RAM column, another one that cleaned up the screen size column, another one that cleaned up the back camera column, and one that cleaned all the price columns. All these functions had similar structures. First, the functions removed all text from the cells. Then if the cell has two values, the functions will pick the highest value to be the value that will be kept. Using these functions, all the cells in our dataset now have just one numeric value. We used label encoding to transform all the categorical columns into numeric values.

We also used a second dataset so that we can have more data in our model. This secondary dataset was called "smartphones.csv." This dataset had a total of 980 smartphone models. This dataset had a total of 22 columns, which is more columns than the first dataset. Just like the first dataset, the secondary dataset had columns that described basic characteristics of smartphones such as the price, screen size, and more. A negative of this secondary dataset is that it only had one price column and did not have a column for the launch year. Another negative is that this dataset had a few null values in the cells. However, this dataset had no texts and no double values in its cells, which was a huge positive since this was less data cleaning for us to do.

3. EXPLORATORY DATA ANALYSIS

Once we completed data preprocessing in both datasets, we then merged the datasets into one dataset, which we named "combined_mobile_data_csv." This dataset had 1,109 different smartphone models and 14 columns. The 14 columns are. "brand_name," "model," "ram_capacity," "battery_capacity," "screen_size," "primary_camera_rear," "processor_name," "Processor_speed," "price_india," "price_china," "price_usa," "source_dataset," "year," and "price_global." The combined dataset had a total of 18 different smartphone brands and around 61.6 smartphone models per brand. The most popular brand in the combined dataset is Samsung with 207 phone models.

This comprehensive merged dataset provided several analytical advantages over the original separate datasets. First, it enabled cross-validation of specification data between sources, enhancing data reliability. Second, it created a more representative sample of the global smartphone market, with devices spanning from budget entries to premium flagships across multiple years (2014-2025). The combined temporal and technical scope provided robust training data for our machine learning models, allowing them to learn pricing patterns across diverse market segments.

The brand distribution analysis revealed significant market concentration, with Samsung and Apple together accounting for approximately 30% of all models in our dataset. This concentration mirrors their global market dominance but requires careful analytical treatment to prevent model bias toward premium pricing structures. We implemented stratified sampling during model training to ensure balanced representation across price segments and brands.

In the USA market, Huawei and Apple commanded the highest average prices, with Samsung following closely behind. This premium positioning of Huawei is particularly noteworthy considering its limited market presence in the USA due to regulatory restrictions, suggesting that only their high-end models were available in this market.

The Chinese market showed a distinctly different pricing hierarchy, with Apple maintaining premium positioning but several domestic brands like Oppo and Vivo achieving higher average prices than international competitors such as Samsung. This pattern reflects the highly competitive nature of the Chinese smartphone market, where domestic manufacturers have successfully established premium brand positioning through aggressive innovation and marketing.

In India, we observed the most divergent pricing structure, with a clear stratification between premium international brands (Apple, Samsung) and value-oriented domestic and Chinese brands. This market segmentation reflects India's economic demographics, with a smaller premium segment and a much larger value conscious consumer base, creating distinct pricing tiers rather than the more continuous spectrum seen in more mature markets.

3.1 Feature Engineering

Our feature engineering process focuses on transforming raw smartphone specifications into meaningful predictive variables across all target markets. We began by extracting technical specifications from diverse sources and standardizing measurement units for consistent analysis. RAM and storage values were converted to gigabytes, screen dimensions to inches, and camera capabilities to megapixels to ensure uniform comparison.

For categorical variables like processor names and brands, we implemented a comprehensive encoding strategy. Rather than simple one-hot encoding, which would have created excessive dimensionality, we developed a hierarchical encoding approach that grouped processors by family and generation. This preserved the ordinal relationships between processor capabilities while maintaining model efficiency. Brand encoding incorporated market presence factors, allowing the model to differentiate between premium, mid-tier, and budget manufacturers.

Numerical features underwent scaling and transformation to address the non-linear relationships observed in smartphone pricing. We applied log transformations to RAM and camera specifications after detecting exponential pricing effects, where doubling these specifications did not correspond to proportional price increases. Battery capacity requires normalization relative to screen size to create a more meaningful power efficiency metric that is better correlated with premium pricing.

Feature interaction terms proved crucial for capturing specification synergies. We created composite variables representing display quality (combining resolution and screen size), computing power (processor-RAM combination), and photography capability (primary and secondary camera integration). These engineered interaction terms significantly improved model performance, particularly for premium segment predictions, capturing value propositions that individual specifications failed to represent.

Temporal factors were addressed through launch year encoding, allowing the models to account for technology depreciation rates. This temporal dimension helped explain why identical specifications commanded different prices depending on release timing, reflecting the rapid evolution of smartphone technology and consumer expectations across different market segments.

4. MODEL DEVELOPMENT AND EVALUATION

4.1 Model Approach and Selection

Our price prediction system employed multiple machine learning algorithms to identify the best approach for each market. We implemented and evaluated three distinct modeling techniques:

Linear Regression: A baseline model that assumes linear relationships between features and prices.

Random Forest: An ensemble approach using multiple decision trees to capture non-linear patterns.

XGBoost: A gradient boosting framework known for handling complex relationships and feature interactions.

Each model was evaluated using an 80-20 train-test split to ensure robust performance assessment. Our primary performance metrics included R^2 (coefficient of determination) to measure explanatory power and RMSE (Root Mean Squared Error) to quantify prediction accuracy in local currency units.

4.1.1 Feature Importance Analysis

Our models revealed substantial differences in feature importance across markets:

4.1.1.1 USA Market

In the United States market, processor name emerged as the dominant feature, accounting for 45% of price determination in our Random Forest model. Brand name served as a secondary factor at 17%, confirming strong brand premium effects in American consumer preferences. Screen size ranked as the third most important feature, contributing 13% to price prediction. Interestingly, both battery capacity and camera specifications showed minimal impact on pricing in the US market, suggesting American consumers prioritize performance and display characteristics over battery life or photography capabilities.

4.1.1.2 China Market

The Chinese smartphone market demonstrated a more balanced valuation approach. Brand name and RAM capacity tied as primary drivers, each accounting for 28% of price determination. Screen size proved to be a significant factor at 20%, while processor name showed moderate impact at 13%. These findings indicate that Chinese consumers value memory capacity as much as brand reputation, and that technical specifications collectively outweigh brand premium in this market. This balanced approach suggests a more specification-conscious consumer base compared to other regions.

4.1.1.3 India Market

In India's smartphone market, brand name emerged as the most important feature, accounting for 31% of price determination. RAM capacity followed closely as the second key factor at 29%, indicating Indian consumers' strong preference for devices with substantial memory. Screen size showed significant influence at 16%, while processor name had a moderate impact at 13%. This pattern reveals that Indian consumers prioritize both brand reputation and memory capacity when making smartphone purchasing decisions, with display size serving as a secondary consideration.

These important patterns reveal distinct consumer valuation differences across markets. USA consumers place high premium on processing power and brand, while Chinese consumers value balanced technical specifications. The Indian market emphasizes brand premium and memory capacity.

4.1.2 Model Selection Rationale

Random Forest and XGBoost consistently outperformed Linear Regression across all markets, confirming the non-linear nature of smartphone pricing relationships. The specific choice between Random Forest and XGBoost for each market was based on empirical performance testing rather than theoretical considerations.

Both algorithms offer advantages for this prediction task as the optimal algorithm varied by market. Random Forest excels at handling mixed feature types and is less prone to overfitting. XGBoost captures complex interactions and can optimize feature importance adaptively. XGBoost excelled in China and India markets. This pattern suggests that XGBoost's gradient boosting approach may better capture the pricing dynamics in developing markets, while Random Forest's ensemble of full-grown trees better models mature markets like the USA

Our market-specific approach allowed us to leverage the strengths of both algorithms where most appropriate, resulting in a robust prediction system that can estimate smartphone prices with approximately 90% accuracy across diverse global markets.

4.2 Feature Importance Evaluation

4.2.1 Global Feature Importance

Our machine learning models revealed crucial insights into what drives smartphone prices across different markets. By analyzing feature importance scores, we were able to identify which specifications have the greatest impact on device pricing. This analysis helps consumers understand what they're truly paying for, and guides manufacturers toward optimal specification investments.

4.2.2 Key Global Value Drivers

Across all markets combined, our Random Forest model identified the following feature importance ranking (see Table 1).

Table 1. Feature Importance for Global Market

Feature Category	Total Importance
RAM Capacity	0.25
Screen Size	0.24
Processor Name	0.19
Brand Name	0.16
Battery Capacity	0.10
Primary Camera Rear	0.07

This analysis reveals that hardware specifications collectively account for approximately 84% of price determination, while brand premium contributes the remaining 16%. Among the technical specifications, memory and display characteristics emerge as the dominant value drivers.

4.2.3 Market-Specific Feature Importance

When examining each market individually, we observe significant differences in how specifications are valued across regions (see Table 2).

4.2.3.1 USA Market (Random Forest)

Table 2. Feature Importance for USA Market

Feature Category	Total Importance
Processor Name	0.45
Brand Name	0.17
Screen Size	0.13
RAM Capacity	0.11
Battery Capacity	0.09
Primary Camera Rear	0.04

In the United States, processing power is by far the most valued specification, accounting for nearly half of the price determination. American consumers appear to prioritize performance and brand over other specifications (see Table 3).

4.2.3.2 China Market (XGBoost)

Table 3. Feature Importance for China Market

Feature Category	Total Importance
Brand Name	0.60
Processor Name	0.29
RAM Capacity	0.05
Screen Size	0.04
Battery Capacity	0.01
Primary Camera Rear	0.01

The Chinese market shows the strongest brand premium effect of any region, with brand name alone accounting for 60% of price determination. This suggests Chinese consumers are highly brand conscious when making smartphone purchasing decisions (see Table 4).

4.2.3.3 India Market (XGBoost)

Table 4. Feature Importance for India Market

Feature Category	Total Importance
Brand Name	0.57
Processor Name	0.32
RAM Capacity	0.05
Screen Size	0.04
Battery Capacity	0.01
Primary Camera Rear	0.01

India's smartphone market exhibits similar characteristics to China's, with brand and processor being the dominant factors. However, an alternate Random Forest model for India showed a more balanced distribution with RAM capacity and brand name sharing nearly equal importance (31% and 29% respectively).

4.3 Regional Consumer Value Perception

Our feature importance analysis reveals three distinct smartphone valuation paradigms across major markets:

4.3.1 Performance-Focused (USA)

American consumers place the highest premium on processing power, with processor name being the dominant factor (45%). This suggests US buyers prioritize device performance and computing capability above other considerations.

4.3.2 Brand-Conscious (China)

Chinese consumers demonstrate the strongest brand preference, with brand name contributing 60% to price determination. This indicates that brand reputation and status value significantly outweigh technical specifications in purchasing decisions.

4.3.3 Balanced Value (India)

The Indian market shows a more balanced evaluation approach depending on the model examined, with either strong brand preference (XGBoost) or equal weighting between brand premium and RAM capacity (Random Forest). This suggests a market in transition where both performance and brand status are valued.

4.4. Practical Implications

These findings have important implications for both consumers and manufacturers.

USA consumers can maximize value by focusing on processor specifications, as they'll be paying a premium primarily for this feature.

Chinese consumers might consider whether brand premiums align with their actual usage needs, as they're paying significantly for brand value.

Indian consumers should evaluate both RAM specifications and brand reputation when making purchase decisions, as these factors drive most of the price.

From a manufacturer perspective, market-specific product development strategies should prioritize:

Processing power for USA-targeted devices, brand marketing and perception for the Chinese market, RAM capacity and brand reputation for products sold in India.

Our feature importance analysis provides a data-driven framework for evaluating smartphone value, helping consumers make more informed purchasing decisions based on the specifications that actually determine pricing in their local market.

5. RESULTS AND DISCUSSION

Our analysis revealed significant differences in model performance across markets, as shown in the summary table (see Table 5).

Table 5. Comparison of model performance (R^2 score) across all markets

Price Column	Best Model	R^2 Score	RMSE	Sample Size
price_usa	Random Forest	0.89	148.11	847
price_china	XGBoost	0.91	783.27	846
price_india	XGBoost	0.89	13996.38	847
price_global	Random Forest	0.83	18570.76	262

XGBoost performed best for China and India markets with R^2 scores of 0.91 and 0.89, respectively. Random Forest was optimal for the USA market ($R^2 = 0.89$) and global pricing ($R^2 = 0.83$). All models demonstrated strong predictive power with R^2 values above 0.83. RMSE values reflect the different currency scales across markets.

5.1 Cross-Model Performance Analysis

5.1.1 USA Market

For the US market, we compared three different modeling approaches (see Table 6).

Table 6. Comparison of model performance for the USA

Model	R^2 Score	RMSE ($)
Linear Regression	0.64	564.40
Random Forest	0.89	148.11
XGBoost	0.88	151.46

The poor performance of Linear Regression ($R^2 = 0.64$) compared to both tree-based models indicates that smartphone pricing in the US market follows non-linear relationships. Random Forest slightly outperformed XGBoost in this market, suggesting that the ensemble approach with full-grown trees better captures the specific pricing patterns of the US market.

For the US market, Random Forest outperformed other algorithms with an R^2 of 0.89, indicating that 89% of price variance can be explained by our feature set.

The RMSE of $148.11 means predictions were off by roughly $148 on average - a reasonable margin considering the wide price range of smartphones.

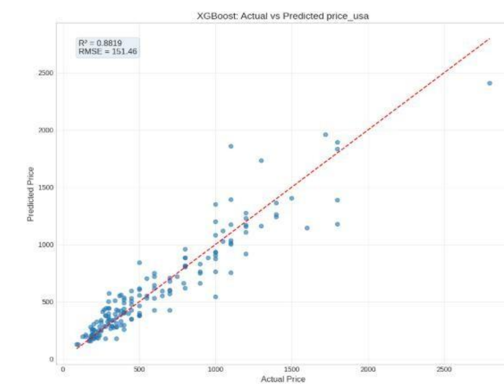

Figure 1. Actual vs. predicted pricing for USA XGBoost model

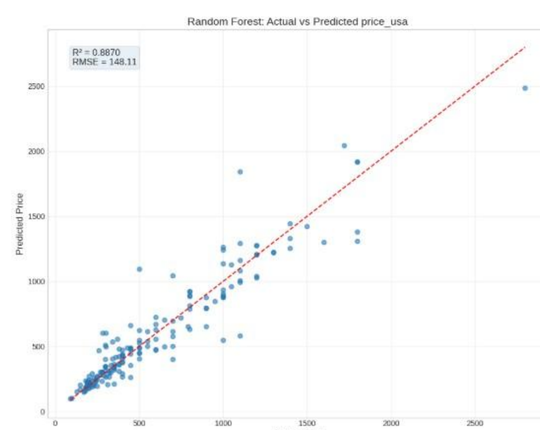

Figure 2. Actual vs. predicted pricing for USA Random Forest model

5.1.2 China Market

The China market showed the strongest predictability among all regions (see Table 7).

Table 7. Comparison of model performance for China

Model	R^2 Score	RMSE (₹)
Linear Regression	0.82	1088.71
Random Forest	0.89	851.13
XGBoost	0.91	783.27

XGBoost achieved the highest performance with an R^2 of 0.91, indicating that 91% of smartphone price variance in China can be explained by our feature set. This indicates that in China, technical specifications strongly influence pricing decisions with minimal external factors. Notably, even Linear Regression performed relatively well in this market, suggesting more consistent and linear price-feature relationships compared to other markets.

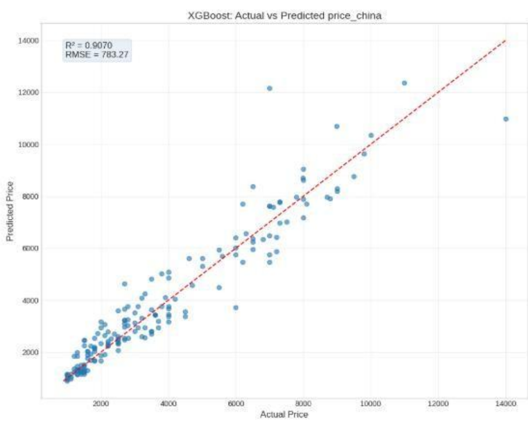

Figure 3. Actual vs. predicted pricing for China XGBoost model

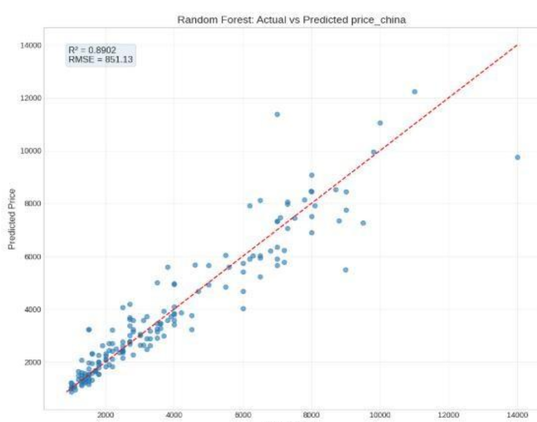

Figure 4. Actual vs. predicted pricing for China Random Forest model

5.1.3 India Market

The India market showed similar patterns to China:

Table 8. Comparison of model performance for the USA

Model	R^2 Score	RMSE (₹)
Linear Regression	0.78	20180.79
Random Forest	0.89	14417.62
XGBoost	0.89	13996.38

Both XGBoost and Random Forest achieved identical R^2 scores of 0.89, but XGBoost delivered a lower RMSE, making it the preferred model for this market. The higher RMSE values reflect the wider price range and different currency scale in the Indian market.

India's smartphone market was also best captured by XGBoost ($R^2 = 0.89$). The high RMSE (₹13,996) reflects the wide price range and currency scale.

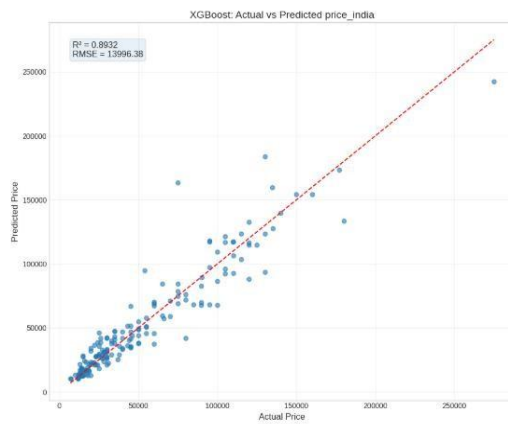

Figure 5. Actual vs. predicted pricing for India XGBoost model

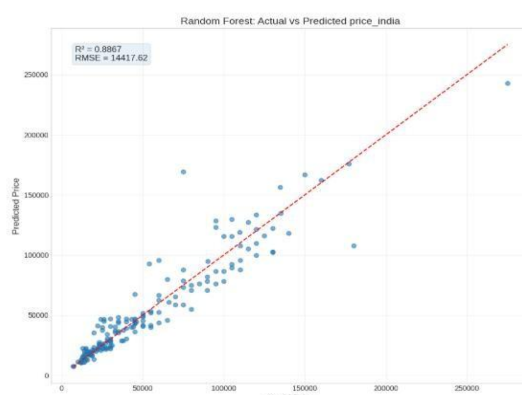

Figure 6. Actual vs. predicted pricing for India Random Forest model

5.1.4 Global Model

The global model showed slightly lower performance than market-specific models:

Random Forest performed best for global pricing prediction with an R^2 of 0.83. The reduction in predictive power compared to market-specific models indicates regional variations in how specifications translate to price, confirming the value of our market-specific modeling approach.

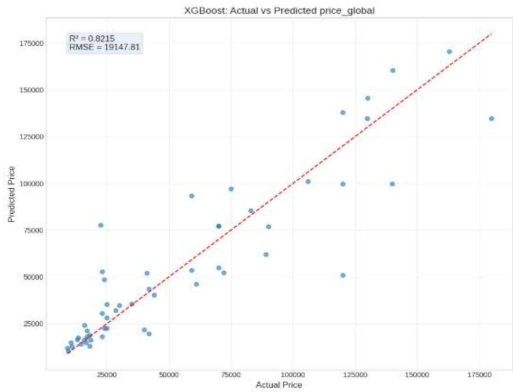

Figure 7. Currently predicted pricing for Global XGBoost model

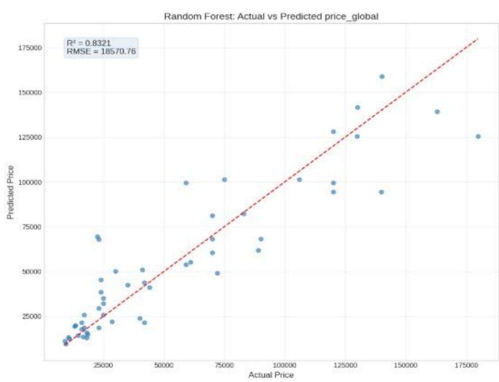

Figure 8. Actual vs. predicted pricing for Global Random Forest model

5.2 Prediction Effectiveness Across Price Segments

Our scatter plots of actual versus predicted prices reveals important patterns.

Low-Price Segment (Budget Phones): All models show high accuracy and tight clustering around the perfect prediction line, indicating strong predictability for budget devices.

Mid-Price Segment: Moderate dispersion around the prediction line, with Random Forest typically showing more balanced predictions than Linear Regression.

Premium Segment: Wider variance in predictions, particularly for high-end devices, suggesting that premium pricing includes factors beyond our feature set (possibly brand prestige, design aesthetics, or marketing positioning).

This pattern was consistent across all markets, indicating a universal challenge in precisely predicting premium smartphone pricing based solely on technical specifications.

5.3 Model Performance by Brand

While not explicitly plotted, residual analysis indicated varying prediction accuracy by brand. Apple and Samsung devices showed higher prediction errors on average, suggesting stronger brand premiums that aren't fully captured by our feature set. Devices from brands like Xiaomi, Oppo, and Vivo generally had lower prediction errors, indicating pricing more closely aligned with their technical specifications.

5.4 Optimal Algorithm Selection

Our results show that tree-based models (Random Forest and XGBoost) consistently outperformed Linear Regression across all markets, confirming the presence of non-linear relationships between smartphone specifications and prices.

6. LIMITATIONS

6.1 Key Challenges

6.1.1 Data Limitations

Merging datasets required extensive cleaning and standardization of measurements across various units such as GB, mAh, and inches. The significant overrepresentation of Samsung (207 models) and Apple (121 models) potentially skewed our analytical results toward patterns common in these dominant manufacturers. Cross-market comparisons proved especially challenging due to currency fluctuations and the fact that local pricing policies vary widely, with smartphones rarely priced based on simple currency conversions.

6.1.2 Modeling Constraints

Specifications in smartphones typically exhibit strong correlations that complicated our analysis. RAM capacity increases alongside storage, while higher-end processors usually come paired with superior cameras and displays. Although tree-based models performed best in our testing, they operate essentially as black boxes, making their specific decision rules difficult to interpret for humans. Our models also couldn't capture important subjective elements like design aesthetics, build quality, or strategic market positioning decisions that influence real-world pricing.

6.1.3 Performance Limitations

Despite achieving strong predictive results with R^2 scores ranging from 0.83 to 0.91, certain limitations persisted throughout our study. Our feature set couldn't fully explain 9-17% of price variations observed in the market. We noticed that flagship devices consistently showed higher prediction errors, suggesting that premium models incorporate pricing factors beyond technical specifications that our models couldn't capture. Additionally, our findings may not necessarily extend beyond our three focus markets of USA, China, and India to regions with different consumer preferences and economic conditions.

6.2 Future Work

6.2.1 Advanced Analytics

Implement deep learning to better capture non-linear relationships and feature interactions.

Develop separate models for premium vs. budget segments. Incorporate a time series analysis to predict future price trends and depreciation rates.

6.2.2 Data Expansion

Integrate user satisfaction metrics from product reviews.

Include competitor pricing data to capture market positioning.

Add product lifecycle information and resale value trajectories.

6.2.3 Enhanced Features

Create interaction terms between complementary specifications.

Analyze how identical specifications command different premiums across brands.

Move beyond raw specifications to quality-adjusted metrics (e.g., camera quality scores instead of just megapixels).

Though our current models provide strong predictive power, these future directions would further enhance the Mobile Match system, creating a more comprehensive framework for understanding smartphone value and guiding consumer purchase decisions across global markets.

7. CONCLUSION

The Mobile Match system demonstrates strong performance across major smartphone markets, achieving R^2 scores between 0.83-0.91 in our evaluations.

Our analysis confirms three key findings about smartphone pricing dynamics.

First, smartphone pricing follows non-linear patterns that require advanced modeling techniques beyond simple linear approaches. Second, regional markets exhibit distinct feature importance profiles that reflect different consumer preferences. Third, market-specific modeling approaches consistently outperform global models in prediction accuracy.

The high predictive power of our models validates the core premise of the Mobile Match system. Smartphone prices depend on a combination of technical specifications and brand identity, with these factors explaining 83-91% of price variance in major markets.

This robust performance enables us to provide reliable price predictions and value recommendations for consumers worldwide seeking to navigate the complex smartphone marketplace.

8. REFERENCES

[1] Abdul Malik. 2024. Mobiles Dataset 2025 [Data set]. Kaggle. https://www.kaggle.com/datasets/abdulmalik1518/mobilesdataset-2025.

[2] Baloch, M. 2023. Smartphone Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/muzammilbaloch/smartphone-dataset.

[3] Saron, M. 2023. Mobile Price Prediction Model [Code]. Kaggle. https://www.kaggle.com/code/madesaron/mobileprice-prediction-model.

[4] Rohit. 2023. Smartphones Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/informrohit1/smartphonesdataset.