

# Naïve Bayes Optimization for Visual Ergonomics Prediction in Smartphone Display Mode

Rizkha Tegar Pambudi\*, Agung Wibowo

Faculty of Computer and Education, Informatics Engineering, Ngudi Waluyo University, Semarang, Indonesia

Email: <sup>1,\*</sup>rizkhategar@gmail.com, <sup>2</sup>agungwibowo@unw.ac.id

Correspondence Author Email: rizkhategar@gmail.com

Submitted: 07/11/2025; Accepted: 26/12/2025; Published: 26/12/2025

**Abstract**—Smartphone display mode plays a significant role in preserving visual comfort and limiting the danger of digital eye strain due to prolonged use. This work attempts to maximize the effectiveness of the Naïve Bayes algorithm to predict user display mode preferences based on seven visual ergonomics characteristics, including age, gender, duration of usage, dominant time, purpose of use, screen type, and lighting circumstances. Data was acquired through an online poll of 283 respondents using convenience sample procedures. The main challenge faced was the imbalance in the initial data class distribution, with a ratio of 63% dark mode preference to 37% light mode, which could potentially cause prediction bias in the standard model. To tackle this, a dual optimization technique was applied: the Synthetic Minority Over-sampling Technique for Nominal (SMOTEN) to balance the data, along with GridSearchCV to discover the appropriate hyperparameters. The evaluation results indicated a considerable improvement in performance; the baseline model had an accuracy of 68.42% with a light mode recall value of 0.57, while the optimal model had an accuracy of 70.18% with a significant increase in recall to 0.71. This improvement indicates that the integration of SMOTEN significantly decreases majority prejudice, resulting in a more sensitive and trustworthy model as a foundation for the creation of adaptive interfaces directed towards visual health.

**Keywords:** Visual Ergonomics; Dark Mode; Light Mode; Naïve Bayes; Model Optimization; SMOTEN; Adaptive Interface

## 1. INTRODUCTION

Smartphone use is increasing in various daily activities, and this has an impact on the visual health of users [1]. Exposure to blue light and high screen brightness levels cause eye fatigue, especially when users are in low-light conditions [2]. Light and dark display modes were developed to reduce the visual strain caused by excessive screen light. This feature is an important part of visual ergonomics because it can adjust light intensity to the user's environmental conditions [3]. Display modes affect the visual perception of smartphone users. Dark mode can reduce the intensity of light entering the retina in low-light conditions [4]. Light mode still provides good readability when users are in bright environments because the contrast between text and background increases [5]. This condition shows that visual comfort depends on the suitability between the display mode and the actual lighting [6]. Factors such as ambient lighting, interface background color, and color composition affect the user's comfort level [7]. Previous studies have found that differences in interface design on Android and iOS result in varying visual experiences [8].

Variations in user eye conditions also affect comfort because sensitivity to light and contrast differs between individuals. Users with certain visual conditions are more prone to fatigue when the display does not suit their visual needs. These variations in visual ability result in different display mode preferences for each user [9]. Technical aspects of the interface also affect the visual experience. The spatial layout and color mapping of icons can increase visual search speed and navigation accuracy. A disproportionate interface design can increase cognitive load and reduce comfort [10]. Users need time to adapt to contrast changes to reduce the likelihood of eye fatigue [11]. Age and duration of smartphone use also influence display mode preferences. Older users more often choose light mode for readability. Users with long usage durations tend to choose dark mode to reduce glare [12].

Visual preferences are not only influenced by physiological aspects. Aesthetic preferences and emotional factors also play an important role. Users assess display modes based on comfort, suitability of the display to the context, and aesthetic preferences. This condition makes predicting display mode preferences complex because it involves visual, emotional, and contextual dimensions simultaneously [13]. A data-driven approach is needed to comprehensively understand visual ergonomics preference patterns. The Naïve Bayes algorithm can be used to predict display mode choices based on contextual factors such as lighting, duration of use, and screen type [14]. Previous research shows that Naïve Bayes is capable of processing categorical and text data well, so it is often used in sentiment analysis [15]. Another study proves that model accuracy increases after synthetic oversampling is performed on minority classes [16]. Naïve Bayes works efficiently on large data and can handle unbalanced data, but its performance is highly dependent on class proportions [17]. In the field of education, this method shows good predictive performance when processed using appropriate data processing techniques [18].

Class imbalance is a significant problem in Naïve Bayes-based prediction models for display mode preference data. Imbalance causes the model to predict the majority class more often and ignore minority patterns. This condition reduces performance, especially in the recall metric, which is important to ensure that the model is able to recognize rarely occurring class preferences. Synthetic oversampling techniques such as SMOTEN are used to overcome this problem by adding minority class representations based on patterns similar to the original data [16]. The urgency of addressing imbalance is high because visual ergonomics preference data often shows dominance in one display mode.

Users tend to consistently choose a particular mode, resulting in an unbalanced class distribution. This condition makes Naïve Bayes optimization through oversampling a key requirement for improving prediction performance.

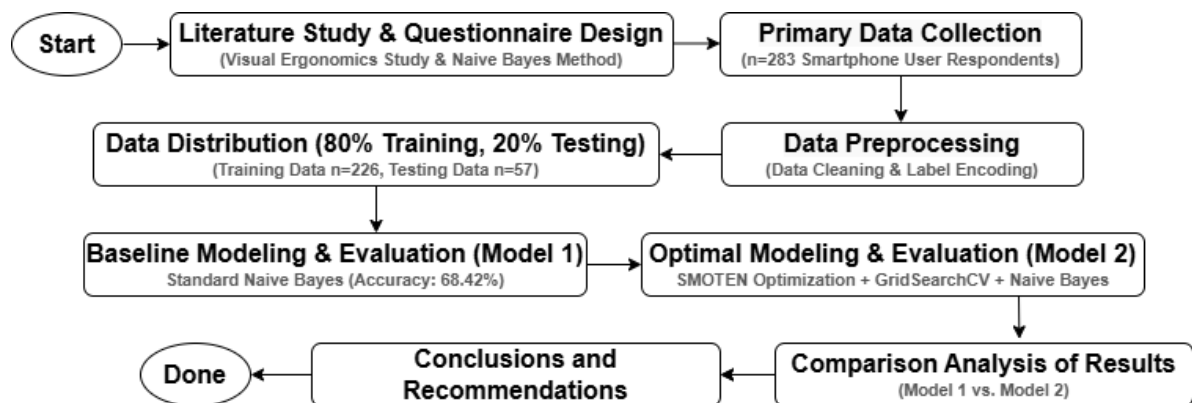
Previous studies have extensively examined the relationship between display modes and visual comfort. However, research combining predictive approaches with data optimization techniques to produce more accurate visual preference models is still limited [14]. Most studies still assess physiological or perceptual aspects without integrating them into adaptive machine learning models [13]. This opens up opportunities to develop predictive models capable of recognizing preference patterns by considering visual and contextual factors simultaneously. This study aims to optimize the Naïve Bayes algorithm to predict display mode preferences based on contextual variables such as lighting, duration of use, and device characteristics. This study also emphasizes class imbalance handling to improve model performance on recall metrics through the application of synthetic oversampling techniques.

The contributions of this research include the development of a more accurate visual ergonomics preference prediction model using Naïve Bayes optimization. This research also provides an approach to handling imbalance in display preference data through synthetic oversampling, which improves the model's ability to recognize patterns in minority classes. This approach forms the basis for an adaptive interface system that can automatically adjust the display mode to improve visual comfort for users.

## 2. RESEARCH METHODOLOGY

### 2.1 Research Stages

The study stages were established systematically to provide a logical and replicable approach, as depicted in Figure 1.



**Figure 1.** Research Process Flow

### 2.2 Data Collection

The initial stage of the research was the collecting of primary data through the dissemination of online questionnaires utilizing the Google Forms platform. The study used a non-probability sampling method, specifically a combination of purposive and convenience sampling. The questionnaire was given to people who satisfied the requirements for inclusion: they were active smartphone users and understood how display modes worked. After this process, we got data from 283 genuine respondents, which was then prepared for analysis.

It's important to recognize the inherent limits of online surveys and convenience sampling. These methods can introduce selection bias, which means the group of people surveyed might be skewed towards those in their working years or those who are comfortable with technology. Therefore, the interpretation of this study's conclusions centers on the specific characteristics of the sample studied, and these findings shouldn't be automatically applied to other smartphone users. Furthermore, the obtained questionnaire data was categorized into eight primary factors. These included seven independent variables, which are traits that might affect one dependent variable, the target. Table 1 provides a detailed description of the operational definitions and measurement methods used for each variable.

**Table 1.** Operational Definitions and Measurement of Variables

Variable	Role	Data Type	Category/Measurement Scale
Age	Features (X1)	Categorical	< 20 Age
			21 – 25 Age
			26 – 30 Age
			> 30 Age
Gender	Features (X2)	Categorical	Male
			Female
Duration of Use	Features (X3)	Categorical	Less than 3 hours
			Between 3 – 6 hours



Variable	Role	Data Type	Category/Measurement Scale
Dominant Time	Features (X4)	Categorical	More than 6 hours
			Morning (6:00 a.m. – 12:00 p.m.)
			Afternoon – Evening (12:00 p.m. – 6:00 p.m.)
			Evening (6:00 p.m. – 12:00 a.m.)
			Early morning (12:00 a.m. – 6:00 a.m.)
Primary Objective	Features (X5)	Categorical	Communication
			Social Media and Entertainment
			Gaming
Screen Type	Features (X6)	Categorical	AMOLED / OLED / Super AMOLED
			IPS LCD / LCD
			Don't know / unsure
Light Conditions	Features (X7)	Categorical	Bright room / Daytime outdoors
			Dimly lit room
			The room is very dark (e.g., before going to sleep)
Preference Mode	Target (y)	Categorical	Dark Mode
			Light Mode

### 2.3 Data Pre-processing

After collecting the main data, it undergoes pre-processing to verify its quality and preparation for the modeling stage. The approach involves cleaning the data to remove extraneous entries, along with transforming features using the Label Encoding method. During this transformation, each unique category inside the feature is changed into distinct integer values, such as 0, 1, 2, and so on. To ensure the accuracy of the data's meaning, coding is applied using a specified method. For ordinal features, like age and how long something has been used, we utilize numbers to show the order of the different levels. In contrast, the Categorical Naive Bayes algorithm helps reduce the possibility of ordering bias when dealing with nominal features that don't have a natural order. Unlike standard methods that treat numbers as continuous data, this algorithm is designed to treat integer inputs as separate categories. It doesn't consider the size or distance between values, which helps to maintain the data structure's integrity during classification.

### 2.4 Data Distribution

After the data was cleaned (n=283), it was separated into two groups: Training Data and Test Data. The split was set at 80% for Training Data, which amounted to 226 samples, and 20% for Test Data, totaling 57 samples. The model was trained using the training data. In contrast, the test data, which the model hadn't seen previously, was kept separate to objectively evaluate the model's performance.

### 2.5 Modeling and Optimization

At this stage, a comparison modeling strategy was used, which involved creating two alternative Naïve Bayes models: the baseline model and the optimal model. The Naïve Bayes Baseline model is the first one, serving as a starting point for comparison. The model uses the standard Categorical Naïve Bayes technique, and it was trained directly on the original training data, which had 226 samples. Because of the class imbalance between the "Dark Mode" and "Light Mode" categories, this model serves as the basic starting point, without any optimization. The second model, Naïve Bayes Optimal, is generated by a series of optimization methods using the same training data. The initial step in the optimization process is using the Synthetic Minority Oversampling Technique for Nominal (SMOTEN). This method seeks to reduce class imbalance by increasing the number of instances in the less-represented class. This creates a more balanced training dataset, resulting in a total of 290 samples. After the training data were balanced, the optimization stage continued using the GridSearchCV approach. Using this method, the best alpha hyperparameter value was systematically found, which led to the most effective configuration for the Naïve Bayes model. After these two optimization steps, we created an Optimal Naïve Bayes Model. This model was trained using balanced training data (n=290) and the best alpha value found by GridSearchCV. The model is predicted to deliver better classification performance than the original baseline model.

### 2.6 Model Evaluation

Evaluating a model is a crucial step in determining how well the algorithm can accurately anticipate data it hasn't seen before. To provide for a fair comparison of performance, this study evaluated the original test data (n = 57) for both the baseline and the best models. The goal of this stage is to assess if the best model demonstrates better performance, especially in identifying minority classes, which are commonly neglected in datasets with uneven distributions. The evaluation used essential metrics: the Confusion Matrix, Accuracy, Precision, Recall, and the F1-Score.

#### a. Confusion Matrix

A confusion matrix is a tabular representation of classification results that describes the relationship between actual labels and model predictions, consisting of four main components, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [19].

This matrix forms the basis for calculating all other metrics. According to Rahman and Irwiensyah [20], the use of a Confusion Matrix can reveal model errors in each class specifically, allowing for a more in-depth evaluation.

b. Accuracy

Accuracy measures the extent to which the model successfully makes correct predictions across the entire dataset. The accuracy formula is given as follows [19], [21]

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

A high accuracy score indicates the model's ability to classify data correctly in general. However, in the context of imbalanced data, accuracy alone is not sufficient to describe the overall performance of the model.

c. Precision

Precision measures how accurately the model identifies positive classes correctly. The precision formula is expressed as follows:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

According to Sasongko and Hilda [21], this metric is very important for cases that are sensitive to positive errors, such as sentiment analysis or spam detection, because it only takes into account correct positive predictions from all positive predictions made by the model.

d. Recall

Recall, also known as Sensitivity or True Positive Rate, measures the extent to which the model is able to find all data that is truly positive. The formula for recall is [22]

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

A high recall value indicates the model's ability to capture all positive data, which is crucial in research with unbalanced data distribution such as public opinion classification [20].

e. F-1 Score

F1-Score is used to balance Precision and Recall, especially when there is a large difference between the two. The F1-Score formula is expressed as follows [19], [21]:

$$F - 1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

This metric provides a fairer view in evaluating model performance on data with an unbalanced class distribution. Fadhlur Rahman and Irwiensyah [20] emphasize that F1-Score is the main indicator in determining classification performance balance.

f. Bayes' Theorem as the Basis of Algorithms

All of the above model evaluations are based on the Naïve Bayes algorithm, which works by applying Bayes' theorem to calculate the probability of a class based on observed features. The basic equation is written as follows:

$$P(C|X) = \frac{P(C|X) \times P(C)}{P(X)} \quad (5)$$

This model calculates the probability of data falling into a particular class by looking at the relationship between features and classes.  $P(C|X)$  shows the probability of data being in class C based on its feature X.  $P(X|C)$  describes the probability of feature X appearing in class C.  $P(C)$  provides information about the initial probability of a class before looking at any features.  $P(X)$  explains the overall probability of the observed features. This theorem allows the model to determine the class with the highest probability. It also supports the calculations used to construct the Confusion Matrix and generate related performance metrics. The same probability structure helps the model evaluate the accuracy of predictions for each class [19].

## 2.7 Comparison of Results and Conclusions

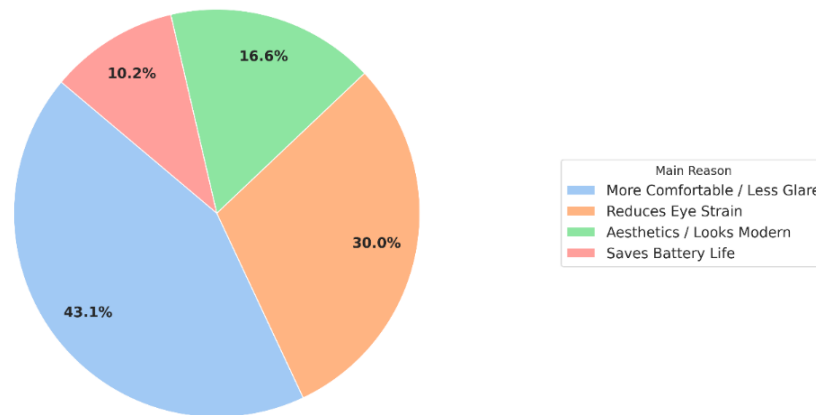
The final stage is Comparison of Results. The performance of the Baseline Model (trained on unbalanced data) and the Optimal Model (trained on balanced and optimized data) is compared directly. This comparative analysis is used to draw research conclusions, namely to determine whether the optimization process (SMOTEN and GridSearchCV) has succeeded in significantly improving the performance of the Naïve Bayes model.

## 3. RESULTS AND DISCUSSION

### 3.1 Data Collection

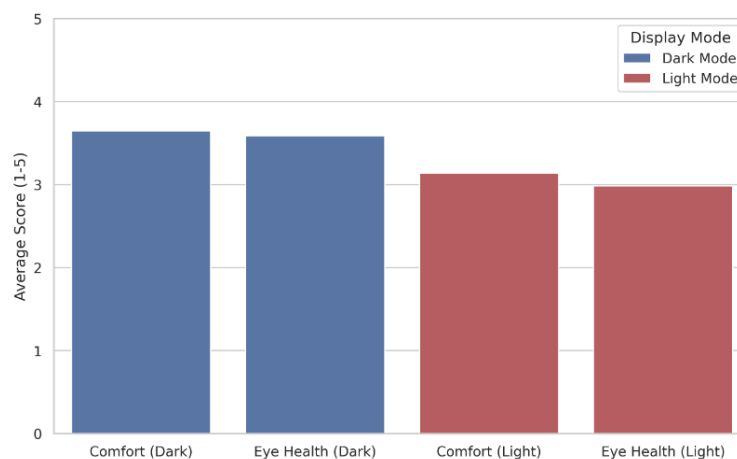
The initial stage of the research process is data collection. The data used in this study is primary data obtained from an online questionnaire (Google Forms) distributed through social media. Data collection was carried out using convenience sampling and purposive sampling techniques targeting active smartphone users. From this process, 283 valid and complete respondent data were collected. The first finding, presented in Figure 2, is an analysis of the main

reasons respondents chose their mode preferences. The results clearly show that visual ergonomics is the dominant driver. The majority of respondents, with a combined total of 99.9 percent, chose their preferences based on the reasons “The display feels more comfortable and is not glaring” (43.1%) or “Reduces eye fatigue/strain” (30%). Other factors such as “Display aesthetics” (16.6%) and “Battery life” (10.2%) were secondary considerations. The discussion of these findings is that the initial hypothesis of the study, which stated visual ergonomics as a key factor, was confirmed by the primary data. Users consciously prioritize eye comfort over aesthetic factors or battery functionality.



**Figure 2.** Distribution of Main Reasons Respondents Chose Display Mode

The second finding is a direct comparison of comfort and health perceptions, visualized in Figure 3. This data provides quantitative justification for the preferences reported in Figure 2. The results show a consistent advantage for Dark Mode in both ergonomic aspects. The average scores (scale of 1-5) for “Dark Mode Comfort” (3.65) and “Dark Mode Eye Health” (3.59) were significantly higher than the scores for “Light Mode Comfort” (3.14) and “Light Mode Eye Health” (2.99). The discussion of these findings is that user perception data reinforces the main reason data. The moderate advantage of Dark Mode (score ~3.6 vs ~3.1) indicates that although Dark Mode is preferred, this preference is not absolute. The clear preference bias in this data also implies that the dataset to be used for modeling is likely to be imbalanced, a challenge that must be addressed in the method implementation stage.



**Figure 3.** Comparison of Average Visual Ergonomics Perception Scores

### 3.2 Data Pre-processing

The raw data from the questionnaire, which included 283 responses, requires preprocessing before machine learning algorithms can be used on it. Converting qualitative data into a format the model can use is a crucial step. Data cleaning is the initial step. This involves stripping away administrative columns that won't be useful for modeling, like timestamps and approval records. Data Transformation is the next step. In this stage, the Likert scale responses, which are still in text format (such "5 (Strongly Agree)"), are cleaned and changed into numerical values, namely the number 5. Label encoding is the third, and most important, procedure. The Naïve Bayes method, which is based on mathematical principles, cannot directly analyze text data, such as strings. Therefore, all categorical data, including age, gender, how long the system is used, the most used time, the main purpose, the type of screen, and the illumination, as well as the target data (mode choice), are translated into numerical values. For example, the age feature is represented by the following mappings: "< 20 Years" is coded as 0, "21-25 Years" as 1, and so on. In this context, "Dark Mode" is represented by the value 0, whereas "Light Mode" is represented by the value 1. Table 2 shows the consequences of this change.

**Table 2.** Example of Pre-Processed Data Results (Label Encoding)

age (text)	dominant time (text)	preference_mode (text)	age (numeric)	dominant time (numeric)	preference_mode (numeric)
21 – 25 Age	Evening (6:00 p.m. – 12:00 a.m.)	Dark Mode	1	2	0
< 20 Age	Afternoon – Evening (12:00 p.m. – 6:00 p.m.)	Light Mode	0	1	1
26 – 30 Age	Morning (6:00 a.m. – 12:00 p.m.)	Dark Mode	2	0	0

### 3.3 Data Sharing

The cleaned dataset, with 283 entries, was divided into two parts: 80% for training (226 entries) and 20% for testing (57 entries). The analysis of the test data shows a class imbalance. The Dark Mode preference is the most common, making up 63% of the data (36 samples out of 57 total). In contrast, the Light Mode preference is less common, at 37% (21 samples out of 57 total). Therefore, the split was done using the stratification method, which helped maintain these proportions consistently in both groups. A prediction model was built using the training data. The test data was then used to evaluate how well the model performed on new data that the system hadn't seen before. Table 3 provides a detailed look at the division's structure.

**Table 3.** Dataset Distribution

Data Type	Proportion	Number of Samples
Training Data	80%	226
Test Data	20%	57
Total	100%	283

### 3.4 Modeling and Optimization

The modeling stage was carried out using two approaches: the Naïve Bayes Baseline Model and the Naïve Bayes Optimal Model. The Baseline Model was built with the original training data (n=226) without class balancing. The Optimal Model was developed through two optimization steps. First, the SMOTEN technique was applied to balance the training data, resulting in 290 samples with a balanced class distribution. Second, the process of searching for the best alpha hyperparameter was carried out with GridSearchCV, and the optimal value found was alpha=0.001.

### 3.5 Model Evaluation

The model evaluation stage aims to assess the performance level of the Naïve Bayes algorithm in predicting user display mode preferences based on test data that the model has never seen before. The evaluation was conducted on two model versions, namely Model 1 (Baseline) and Model 2 (Optimal). The Baseline Model was built using the original training data without class balancing, while the Optimal Model was the result of applying the Synthetic Minority Oversampling Technique for Nominal (SMOTEN) and adjusting the alpha ( $\alpha$ ) hyperparameter using GridSearchCV.

The testing was conducted using 57 test data samples, consisting of 36 Dark Mode user data and 21 Light Mode user data. This evaluation process produced several key metrics, namely Accuracy, Precision, Recall, and F1-Score, all of which were derived from the Confusion Matrix of each model.

The Confusion Matrix serves as the basis for calculating the performance per class (Dark Mode and Light Mode), so that it can be determined how well the model recognizes each category. The values from the Confusion Matrix are then summarized in a Classification Report, which facilitates comparison of the performance between the two models. Thus, this stage not only assesses overall accuracy, but also measures the model's ability to recognize majority and minority classes in a balanced manner.

#### a. Explanation of Metric Calculations

##### 1. Baseline Model

The performance of Model 1 (Baseline) was calculated based on the results of the Confusion Matrix on test data consisting of 57 samples, comprising 36 Dark Mode class data and 21 Light Mode class data. The recall value for Dark Mode was 0.75, resulting in the number of true predictions (TP) for the Dark Mode class being:

$$TP_{dark} = 0.75 \times 36 = 27$$

The overall accuracy of the model was reported to be 68.42%, equivalent to 39 correct predictions out of 57 test data. Thus, the number of correct predictions for the Bright Mode class can be obtained by:

$$TP_{light} = 39 - 27 = 12$$



Thus, the total correct predictions (TP) are 27 for Dark Mode and 12 for Light Mode. Based on this, the accuracy of the Baseline Model is calculated as follows:

$$Accuracy = \frac{TP_{dark} + TP_{light}}{N} = \frac{27 + 12}{57} = \frac{39}{57} = 0.6842 \text{ (68.42\%)}$$

The Baseline Model Confusion Matrix values are presented as follows:

**Table 4.** Baseline Model Confusion Matrix

Actual / Prediction	Dark Mode	Light Mode
Dark Mode (n=36)	27	9
Light Mode (n=21)	9	12

2. Optimal Model

The performance of Model 2 (Optimal) reported in Table 2 is based on the calculation of metrics from the Confusion Matrix (Table 3). Accuracy, which measures the ratio of total correct predictions (TP + TN) to the total test data (N), is calculated as follows:

$$Accuracy = \frac{TP_{dark} + TP_{light}}{N} = \frac{25 + 15}{57} = \frac{40}{57} = 0.7018 \text{ (70.18\%)}$$

The Optimal Model Confusion Matrix values are presented as follows:

**Table 5.** Optimal Model Confusion Matrix

Actual / Prediction	Dark Mode	Light Mode
Dark Mode (n=36)	25	11
Light Mode (n=21)	6	15

3. Comparison of Classification Reports (Baseline vs. Optimal)

The results of the model performance evaluation on the test data are summarized in Table 6. This table compares the Precision, Recall, and F1-Score values between the Baseline Model and the Optimal Model to provide a comprehensive overview of the improvement in classification performance achieved.

**Table 6.** Classification Reports (Baseline vs. Optimal)

Model	Class	Precision	Recall	F1-Score	Support
Model 1 (Baseline)	Dark Mode	0.75	0.75	0.75	36
	Light Mode	0.57	0.57	0.57	21
Accuracy				0.68	57
Model 2 (Optimal)	Dark Mode	0.81	0.69	0.75	36
	Light Mode	0.58	0.71	0.64	21
Accuracy				0.70	57

Model 1 (Baseline):

The precision values generated from the above matrix are:

$$Precision \text{ Dark Mode} = \frac{27}{27+9} = 0.75$$

$$Precision \text{ Light Mode} = \frac{12}{12+9} = 0.57$$

The recall values generated from the above matrix are:

$$Recall \text{ Dark Mode} = \frac{27}{36} = 0.75$$

$$Recall \text{ Light Mode} = \frac{12}{21} = 0.57$$

The F-1 Score values generated from the above matrix are:

$$F - 1 \text{ Score Dark Mode} = 2 \times \frac{0.75 \times 0.75}{0.75 + 0.75} = 0.75$$

$$F - 1 \text{ Score Light Mode} = 2 \times \frac{0.57 \times 0.57}{0.57 + 0.57} = 0.57$$

The Support value shows the amount of actual data in each class, namely:

Dark Mode Support = 36

Light Mode Support = 21

Model 2 (Optimal):

The precision values generated from the above matrix are:



$$\textit{Precision Dark Mode} = \frac{25}{25+6} = 0.81$$

$$\textit{Precision Light Mode} = \frac{15}{15+11} = 0.58$$

The recall values generated from the above matrix are:

$$\textit{Recall Dark Mode} = \frac{25}{36} = 0.69$$

$$\textit{Recall Light Mode} = \frac{15}{21} = 0.71$$

The F-1 Score values generated from the above matrix are:

$$F - 1 \textit{ Score Dark Mode} = 2 \times \frac{0.81 \times 0.69}{0.81 + 0.69} = 0.75$$

$$F - 1 \textit{ Score Light Mode} = 2 \times \frac{0.58 \times 0.71}{0.58 + 0.71} = 0.64$$

The Support value shows the amount of actual data in each class, namely:

Dark Mode Support = 36

Light Mode Support = 21

### 3.6 Discussion of Results

This research endeavors to assess the efficacy of the Naïve Bayes algorithm in forecasting smartphone display mode preferences, grounded in visual ergonomics principles, while also evaluating the influence of model optimization via a combination of Synthetic Minority Over-sampling Technique for Nominal (SMOTEN) and GridSearchCV. The first testing findings reveal that the base model achieves an accuracy of 68.42%. However, it has a low recall of 0.57 for the minority class, which is the light mode. The model's performance suggests that it failed to identify over half of the real user data that preferred light mode. As a result, the model tends to classify the data as belonging to the majority class, which is dark mode. This trend is a common indication of class imbalance. When the data is unevenly distributed, the model tends to focus on the dominant class, which might lead to the minority class's features being ignored.

The research dataset's class imbalance was produced by the large number of participants who preferred dark mode, which made light mode users a smaller group. This situation produces bias in the learning process of Naïve Bayes probability parameters. These parameters are heavily influenced by how often features appear in each class, represented as  $P(X|C)$ . The absence of data representation inside certain classes introduces instability into probability estimates; this can even lead to the generation of probabilities approximating 0 for infrequent feature combinations. Consequently, the model's capacity to identify users who favor light mode is diminished. To solve this challenge, the use of SMOTEN has proven to be an effective approach. Unlike regular SMOTE, which works with numerical data, SMOTEN is specifically built to handle categorical data. This includes screen kind, how long the screen is used, and the main lighting conditions, as used in this study. This method creates new examples for less common classes by looking at how similar the examples are in the original feature space. By including synthetic representations of light mode in the training data, the model has a broader range of patterns to learn from. Therefore, this approach leads in conditional probability estimates that are more stable and representative. After oversampling, the training dataset becomes balanced, which directly improves the model's sensitivity.

In addition to balancing the data, using GridSearchCV to optimize model parameters was crucial for finding the appropriate alpha value for Laplace smoothing. The GridSearchCV process found that the best alpha value was 0.001. This tiny value suggests that the oversampled data distribution is adequately informative and representative, thereby eliminating the need for extensive smoothing to avoid zero probability. Combining data balancing with parameter adjustment proved useful in enhancing how well the model learned. As a result, the optimization improved the model's performance, achieving an overall accuracy of 70.18% and a light mode class recall of 0.71. As a result, the model can now identify 71% of users who favor light mode, a huge gain above its initial ability to identify only 57%. In a predictive system designed for visual comfort, the ability to correctly identify this less common group is regarded more important than just improving overall accuracy. If a model's main strategy is to predict the most common outcome, its accuracy can be misleading. In contrast, a high recall score suggests that the system is better at including and addressing the requirements of user groups that were previously underrepresented.

Further study of the model's probability shows that Age, Dominant Time of Use, and Primary Purpose are the most important factors for identifying user preferences, more so than technical features like Screen Type. The observed probability distributions suggest that users in older age groups are more likely to choose light mode. This could be because their visual systems require different contrast settings, which is likely owing to reduced eye accommodation. In contrast, the amount of time spent using the device at night constantly increases the likelihood of

the dark mode class. This aligns with the visual ergonomics principle, which aims to eliminate glare in low-light environments. The results demonstrate that the Naïve Bayes model successfully identifies behavioral and demographic patterns as important predictors, even though the assumption of feature independence is a theoretical constraint. A careful look at the metric change shows that, although recall improved in the light mode class, it somewhat declined in the dark mode class. However, this was accompanied by a rise in precision. As a result, the model's predictions of dark mode were more conservative, which reduced the number of incorrect predictions of dark mode when it should have been light mode.

The ability to move decision boundaries is beneficial in practice. This is because the most annoying error for users happens when the system doesn't identify the need for a bright mode and instead uses a dark display when it's not suitable. The external validity of these findings also needs to be considered. Using purposive and convenience sampling, we gathered data from 283 participants. This group was mostly similar in terms of their demographics, particularly their age and their comfort with technology. Therefore, we should be careful when applying these findings to a broader group, such as older adults or people new to technology who use devices with varied screen sizes. Moreover, the data's subjective nature, which comes via questionnaire responses, increases the risk of response bias when contrasted to objective measures utilizing biometric sensors. However, this model shows promise for a lightweight implementation on the device side, making it a good first step toward an adaptive interface system. This study shows that the problem of data imbalance in predicting visual ergonomic preferences may be efficiently solved by using the SMOTEN and GridSearchCV methods. The optimization outcomes not only enhance statistical measures but also advance the model's primary objective: to construct an interface system capable of more precisely interpreting and adjusting to users' visual comfort requirements, including crucial elements like age and the temporal context of usage.

## 4. CONCLUSION

The study's results indicate that the Naïve Bayes algorithm is good in predicting preferred smartphone display modes. This suggests that a data-driven approach can be useful for improving visual ergonomics. Data analysis constantly shows that visual comfort and reduced eye strain are the main elements affecting user choices, beyond solely aesthetic concerns. Although the initial model had an accuracy of 68.42%, which was limited by an uneven distribution of classes, using optimization methods, including SMOTEN for categorical data and GridSearchCV, led to a notable improvement, raising the accuracy to 70.18%. The most important outcome was a considerable increase in the recall of the minority class (bright mode), which rose from 0.57 to 0.71. The increased sensitivity has crucial practical effects on interface design. If a system doesn't correctly identify users' light mode preferences, it could lead to displays that don't fit users' visual needs. This could then reduce the overall quality of using the device. Therefore, the improved model offers a better balance in meeting the different demands of users. Considering the constraints inherent in studies dependent on subjective perception data and a homogenous purposive sample, subsequent investigations are warranted to corroborate these results. This validation should incorporate objective data derived from real-time environmental light intensity (lux) sensors. Furthermore, the exploration of alternative algorithms, including Random Forest, is suggested to facilitate the capture of more intricate feature interaction patterns, thereby enhancing model generalization across a wider demographic.

## REFERENCES

- [1] P. Sengsoon and R. Intaruk, "Immediate Effects of Light Mode and Dark Mode Features on Visual Fatigue in Tablet Users," *Int J Environ Res Public Health*, vol. 22, no. 4, Apr. 2025, doi: 10.3390/ijerph22040609.
- [2] A. Shrestha, S. Shrestha, B. Paneru, S. Paudel, and A. Adhikari, "An Exploration of Effects of Dark Mode on University Students: A Human Computer Interface Analysis," arXiv Preprint, vol. abs/2409.10895, 2024, doi: 10.48550/ARXIV.2409.10895.
- [3] P. Tian *et al.*, "Effects of Paradigm Color and Screen Brightness on Visual Fatigue in Light Environment of Night Based on Eye Tracker and EEG Acquisition Equipment," *Sensors*, vol. 22, no. 11, Jun. 2022, doi: 10.3390/s22114082.
- [4] B. Dandumahanti, P. Chittoor, and M. Subramaniam, "Digital Eye Strain Monitoring for One-Hour Smartphone Engagement Through Eye Activity Measurement System," *J Eye Mov Res*, vol. 18, no. 4, p. 34, Aug. 2025, doi: 10.3390/jemr18040034.
- [5] H. P. Huang, M. Wei, H. C. Li, and L. C. Ou, "Visual Comfort of Tablet Devices Under a Wide Range of Ambient Light Levels," *Applied Sciences*, vol. 11, no. 18, Sep. 2021, doi: 10.3390/app11188679.
- [6] Q. Fan, J. Xie, Z. Dong, and Y. Wang, "The Effect of Ambient Illumination and Text Color on Visual Fatigue under Negative Polarity," *Sensors*, vol. 24, no. 11, Jun. 2024, doi: 10.3390/s24113516.
- [7] A. Yang, C. H. Goh, and L. J. Yi, "The Research Into Dark Mode: A Systematic Review Using A Two-Stage Approach And The S-O-R Framework," arXiv Preprint, vol. abs/2311.00000, 2023, arXiv:2311.00000.
- [8] Z. While and A. Sarvghad, "Dark Mode or Light Mode? Exploring the Impact of Contrast Polarity on Visualization Performance Between Age Groups," *IEEE Visualization and Visual Analytics (VIS)*, Oct. 2024, doi: 10.1109/VIS55277.2024.00050.
- [9] Abd. C. Fauzan and K. Hikmah, "Implementasi Algoritma Naive Bayes Dalam Analisis Polarisasi Opini Masyarakat Terkait Vaksin Covid-19," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 2, pp. 122–128, Jul. 2022, doi: 10.36341/rabit.v7i2.2403.



- [10] M. Zhang, Y. Gong, R. Deng, and S. Zhang, “The Effect Of Color Coding And Layout Coding On Users’ Visual Search On Mobile Map Navigation Icons,” *Front Psychol*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.1040533.
- [11] D. Safitri, S. S. Hilabi, and F. Nurapriani, “Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 8, no. 1, pp. 75–81, Jan. 2023, doi: 10.36341/rabit.v8i1.3009.
- [12] S. Ma, C. Chunyang, H. Khalajzadeh, and J. Grundy, “A First Look at Dark Mode in Real-World Android Apps,” *ACM Transactions on Software Engineering and Methodology*, vol. 33, no. 1, pp. 1–9, 2023, doi: 10.1145/3604607.
- [13] M. R. Atsani, I. A. Mukaromah, and M. H. C. Anugerah, “The Role of Color in User Experience: A Systematic Literature Study of User Preferences for Dark and Light Mode,” *Transactions on Informatics and Data Science*, vol. 2, no. 1, pp. 41–54, Jun. 2025, doi: 10.24090/tids.v2i1.13903.
- [14] F. J. Pathari, Y. Nielsen, L. I. Andersen, and G. Marentakis, “Dark Vs. Light Mode On Smartphones: Effects On Eye Fatigue,” *arXiv Preprint*, vol. abs/2409.10895, 2024, arXiv:2409.10895.
- [15] L. M. Ikhlas, V. P. Rantung, and G. C. Rorimpandey, “Penerapan Algoritma Naive Bayes Classifier dalam Menganalisis Sentimen pada Produk Smartphone Tipe Terbaru,” *Journal of Informatics, Business, Education, and Innovation Technology*, vol. 6, no. 1, 2025, doi: (DOI belum tersedia pada laman jurnal).
- [16] M. Fazri and A. Voutama, “Analisis Sentimen Publik Terhadap Danantara di Media Sosial X Menggunakan NLP dan Pembelajaran Mesin,” *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 9, no. 1, p. 197, Jul. 2025, doi: 10.35145/joisie.v9i1.4924.
- [17] R. Herdian Saputra and R. Randy Suryono, “Perbandingan Algoritma SVM, Random Forest, dan Naive Bayes Terhadap Kasus Scam di Media Sosial Twitter,” *Technology and Science (BITS)*, vol. 7, no. 2, 2025, doi: 10.47065/bits.v7i2.7236.
- [18] A. Wibowo and A. Rohman, “Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ,” *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 12, no. 2, p. 104, Dec. 2022, doi: 10.36448/expert.v12i2.2810.
- [19] Q. Aini, M. N. F. Hidayat, and A. Tholib, “Analisis Sentimen Hasil Pemilu (Quick Count) Calon Presiden dan Wakil Presiden 2024 di Media Sosial Media X Menggunakan Metode Bidirectional Long Short-Term Memory (BiLSTM),” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 3, pp. 690–700, May 2024, doi: 10.47065/josyc.v5i3.5223.
- [20] R. F. Rahman and F. Irwiensyah, “Analisis Sentimen Ulasan Pengguna Aplikasi Alibaba.Com pada Google Playstore Menggunakan Naive Bayes,” *Journal of Computer System and Informatics (JoSYC)*, vol. 6, no. 1, pp. 64–72, Nov. 2024, doi: 10.47065/josyc.v6i1.6132.
- [21] K. H. Sasongko and A. M. Hilda, “Analisis Sentimen Terhadap Film Dirty Vote Pada Media Sosial X dan YouTube Dengan Algoritma Naive Bayes dan SVM,” *Journal of Computer System and Informatics (JoSYC)*, vol. 6, no. 1, pp. 73–85, Nov. 2024, doi: 10.47065/josyc.v6i1.6150.
- [22] A. Suarisman, A. Nazir, F. Syafria, and L. Afriyanti, “Perbandingan Jarak Metrik pada Klasifikasi Jamur Beracun Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 10–19, Nov. 2023, doi: 10.47065/josyc.v5i1.4511.