## Capstone-Project-1:Assessing Impact of Digital Lens Usage on Eye Dryness using Schirmer's Effect

PROJECT REPORT

**Data Science with Python Programming**

**INDUSTRIAL PROJECT BASED LEARNING**



**Department of Computer Science and Engineering**

**Accredited by NBA**

**Geethanjali College of Engineering and Technology**

**(UGC Autonomous)**

(Affiliated to J.N.T.U.H, Approved by AICTE, New Delhi)

Cheeryal (V), Keesara (M), Medchal.Dist.-501 301.

**By Team -4**

**Team Details :**

Shivanoor Vignesh 21R11A0597

G.SandhyaRani 21R11A0569

R.Dhathri 21R11A0545

Hasini 21R11A0518

A.Pooja 21R11A0501

PSBN Sriya 21R11A0595

N Sai Charan 21R11A0588

MD.Akbar Ali 21R11A0586

# ABSTRACT

# Digital eye strain (DES) has emerged as a significant concern in the modern era due to the widespread use of digital electronic devices such as computers, smartphones, and tablets. This condition, characterized by a range of ocular and visual symptoms, has become more prevalent with the shift towards remote work and digital learning, particularly exacerbated by the COVID-19 pandemic. With the prolonged use of digital screens becoming a norm, it is imperative to understand the risk factors associated with DES, especially among undergraduates who heavily rely on these devices for academic and recreational purposes.

# This project aims to address the problem of assessing risk factors related to prolonged use of digital screens on the eyes among undergraduates. Through a comprehensive market survey and analysis of primary data collected, the project explores, visualizes, and models the dataset to develop a ready-to-use product on a web API platform. Utilizing tools such as Anaconda Python Jupyter for EDA, data cleaning, and model building, the project identifies and evaluates the best model for predicting and understanding DES risk factors.

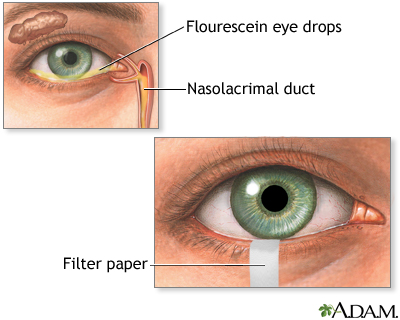
# The outcomes of the project include insights into the prevalence and impact of digital eye strain among undergraduates, as well as recommendations for mitigating its effects. By building a web application using the Flask API framework, the project provides a user-friendly platform for accessing and utilizing the developed model for assessing DES risk factors.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter No** | **Chapter Name** | **Page No** |
|  | **ABSTRACT** | 2 |
| **1** | **INTRODUCTION** | 4 |
| **2** | **LITERATURE SURVEY** | 5 |
| **3** | **PROBLEM STATEMENT** | 6 |
| **4** | **OBJECTIVES** | 7 |
| **5** | **METHODOLOGY** | 8 |
|  | **5.1 Data Source** | 8 |
|  | **5.2 Exploratory Data Analysis** | 10 |
|  | 5.2.1 Slicing and Dicing | 12 |
|  | 5.2.2 Checking for Data Consistency | 12 |
|  | 5.2.3 Data Visualization | 15 |
| **6** | **ALGORITHMS** | 26 |
| **7** | **IMPLEMENTATION** | 26 |
|  | 7.1 KNearNeighbour | 26 |
|  | 7.2 RandomForest Classifier | 27 |
|  | 7.3 Decision Tree Classifier | 28 |
| **8** | **Final Implementation** | 30 |
| **9** | **Final Observation** | 33 |
|  | **Conclusion** | 33 |

# INTRODUCTION

The Schirmer's test is a fundamental diagnostic tool used in ophthalmology to evaluate tear production, aiding in the diagnosis of conditions such as dry eye syndrome. This test involves placing a strip of filter paper inside the lower eyelid to measure the amount of moisture produced by the eyes over a specified time period. It is particularly useful when a person presents with symptoms such as dryness or excessive watering of the eyes, which could indicate underlying ocular health issues. While the test has been in use since 1903, advancements in diagnostic techniques continue to refine our understanding of tear production and dry eye conditions.



Dry eyes may result from:

• Aging

• Swelling or inflammation of the eyelids (blepharitis)

• Climate changes

• Corneal ulcers and infections

• Eye infections (for example, conjunctivitis)

• Sjögren syndrome

• Vitamin A deficiency

• Risks

Dry eye syndrome is a common condition characterized by insufficient tear production or poor tear quality, leading to symptoms such as eye discomfort, redness, itching, and blurred vision. The Schirmer's test provides valuable information about tear film dynamics by measuring the amount of tears produced over a specified time period. Not detecting Schirmer's eye abnormalities and addressing dry eye syndrome in its early stages can result in worsening symptoms, ocular surface damage, visual impairment, and increased susceptibility to infections and complications. Early diagnosis and appropriate management are essential for preserving ocular health, maintaining visual function, and improving patients' quality of life.

# LITERATURE SURVEY

Digital eye strain (DES) is a common condition that has become even more prevalent in recent times due to the increased use of digital devices. According to a study by Rosenfield et al. (2014), the prevalence of DES has been estimated to be between 25% and 93%, with the wide range attributed to differences in study populations, measurement methods, and definitions of DES. Another study by Lin et al. (2014) found that the prevalence of DES among university students was 89.8%, with female students being more likely to experience symptoms than male students.

Several studies have investigated the risk factors associated with DES. A study by Sheppard and Wolffsohn (2013) identified factors such as prolonged screen time, inadequate blinking, and poor lighting conditions as contributing to DES. Similarly, a study by Tears and Eye Health Study Group (2017) found that prolonged visual display terminal (VDT) use, age, and female gender were significant risk factors for DES. The study also found that individuals who experience DES are more likely to have dry eye symptoms, which is consistent with the findings of Chen et al. (2018), who reported that DES was significantly associated with dry eye symptoms.

The Schirmer's test is a common diagnostic test used to assess tear production and is often used in the diagnosis of dry eye syndrome. However, the reliability and validity of the test have been questioned. According to a study by Begley et al. (2008), the Schirmer's test has low sensitivity and specificity, particularly in individuals with mild to moderate dry eye symptoms. The study suggested that other tests, such as tear film break-up time and ocular surface staining, may be more helpful in diagnosing dry eye syndrome.

In contrast, a study by Goto et al. (2016) found that the Schirmer's test was a useful tool for diagnosing dry eye syndrome, particularly in patients with mild to moderate symptoms. The study also found that combining the Schirmer's test with other diagnostic tests, such as tear osmolarity and matrix metalloproteinase-9 (MMP-9) levels, improved the accuracy of the diagnosis.

In addition to the Schirmer's test, other diagnostic tests have been developed to assess dry eye syndrome. For example, a study by K conjunctival impression cytology, which involves taking a sample of the conjunctiva and examining it under a microscope, was found to be a useful tool for diagnosing dry eye syndrome (Baudouin et al., 2008). Another study by Dogru et al. (2014) found that tear osmolarity was a reliable and accurate measure of dry eye severity.

# PROBLEM STATEMENT

In today's digital age, the widespread use of electronic devices has become an integral part of daily life, impacting eye health in various ways. To fully grasp the effects of digital screen usage on eye health and associated symptoms, it is essential to analyze a multitude of factors. These include age, duration of screen time, online platforms, nature of activities, screen illumination, years of exposure, daily screen hours, types of devices used, distance from the screen, nighttime usage, blinking frequency, difficulty in focusing, frequency and severity of complaints, observed ocular symptoms, and specific eye examination results. By meticulously examining these variables, we can uncover patterns and correlations that inform the development of effective strategies for maintaining optimal eye health in the digital age. This understanding is crucial for promoting healthy screen habits and mitigating potential risks associated with digital screen usage among individuals, particularly undergraduates who may be more susceptible to the impacts of prolonged screen time.

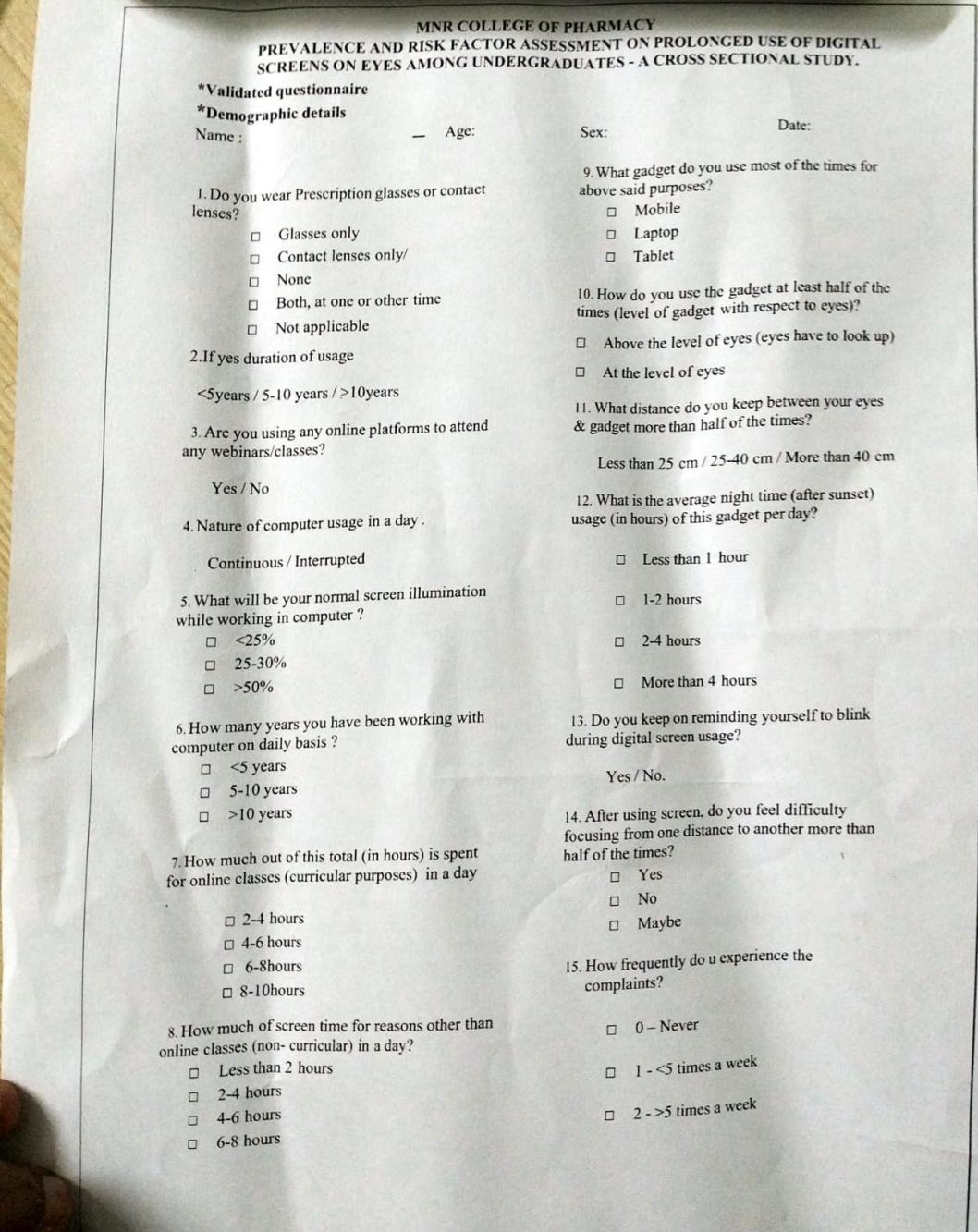
# 4.OBJECTIVES

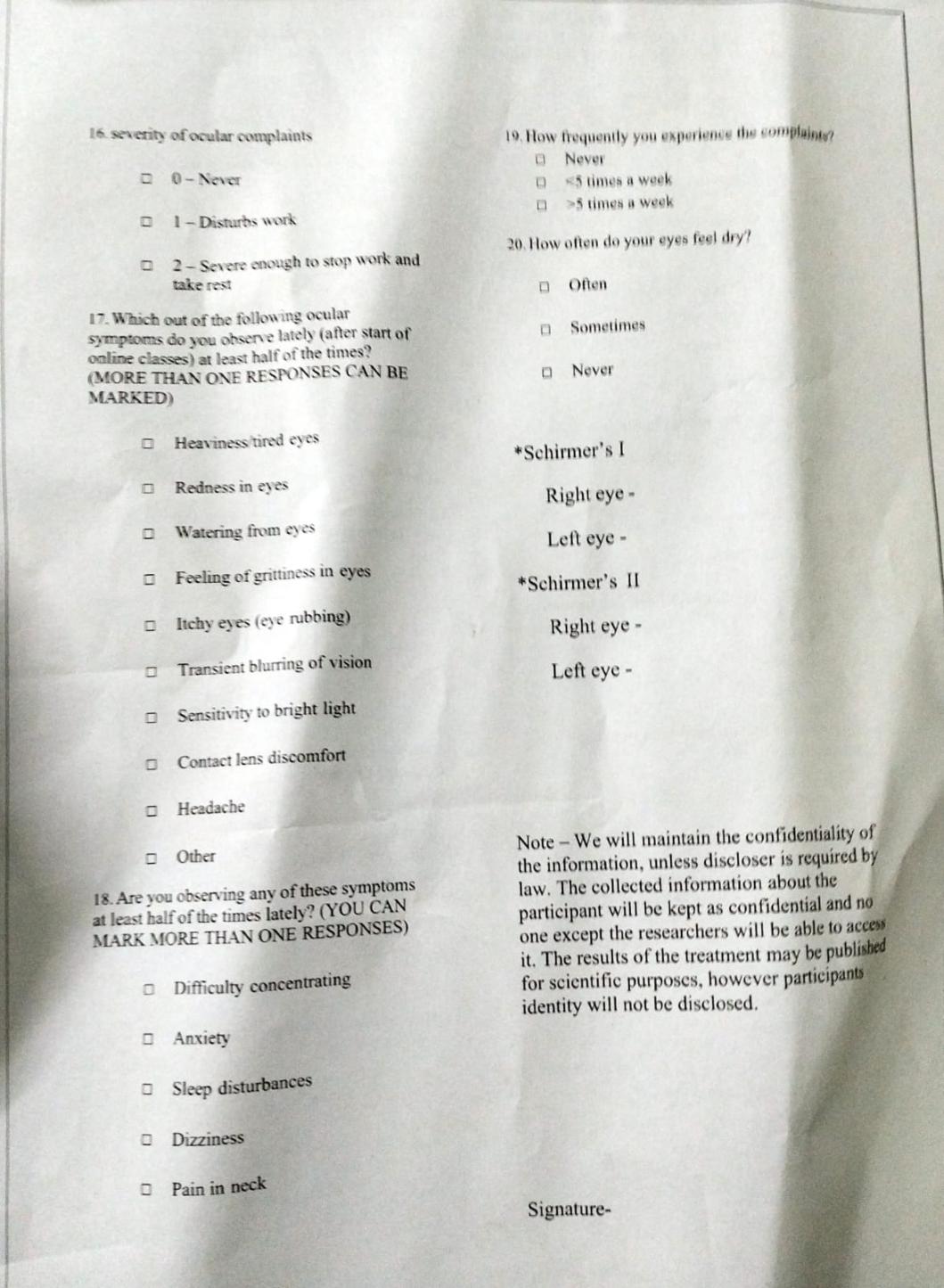
* Develop a predictive model specifically tailored for detecting abnormalities in Schirmer's eye test results.
* Classify Schirmer's test outcomes into normal or abnormal categories using various machine learning classification techniques.
* Compare and evaluate different machine learning classification algorithms to identify the most effective model for accurately detecting abnormalities in Schirmer's test results.
* Present key performance metrics, such as accuracy, precision, recall, and F1-score, to comprehensively assess the predictive performance of the developed models.
* Explore and propose novel features or model architectures that have the potential to enhance the accuracy and reliability of predicting abnormalities in Schirmer's test results.

**5.METHODOLOGY**

**5.1 Data Source**

The dataset is procured from mohan pulipaka sir and survey from which represents the values of the columns is also given





* **Brief description of the data source**

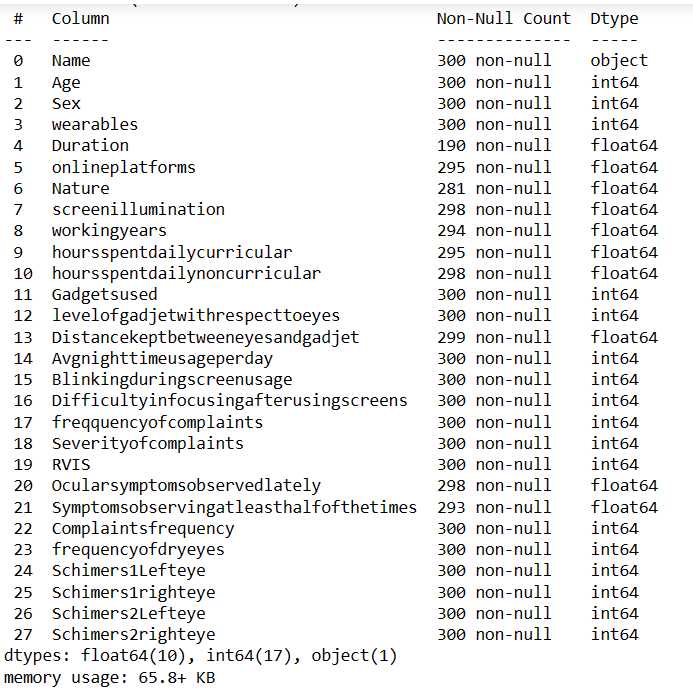
|  |  |  |
| --- | --- | --- |
| **S.no** | **Feature name** | **Description** |
| **1.** | **Name** | Unique identifier for each person |
| **2.** | **Age** | Age of a person |
| **3.** | **Sex** | Gender of a person |
| **4.** | **Wearables** | Presence of wearable technology devices |
| **5.** | **Duration** | Time duration of activity or usage |
| **6.** | **online-platforms** | Usage of online platforms or applications |
| **7.** | **Nature** | Engagement with outdoor or natural environments |
| **8.** | **screenillumination** | Intensity of screen brightness |
| **9.** | **workingyears** | Number of years spent in employment |
| **10.** | **hoursspentdailycurricular** | Time spent on academic-related activities per day |
| **11.** | **hoursspentdailynoncurricular** | Time spent on non-academic activities per day |
| **12.** | **Gadgetsused** | Usage of electronic devices or gadgets |
| **13.** | **levelofgadjetwithrespecttoeyes** | Positioning of gadgets relative to the eyes |
| **14.** | **Distancekeptbetweeneyesandgadjet** | Proximity of gadgets to the eyes |
| **15.** | **Avgnighttimeusageperday** | Average duration of nighttime screen usage |
| **16.** | **Blinkingduringscreenusage** | Frequency of blinking while using screens |
| **17.** | **Difficultyinfocusingafterusingscreens** | Challenges in maintaining focus post-screen usage |
| **18.** | **freqquencyofcomplaints** | Frequency of reported issues or discomfort |
| **19.** | **Severityofcomplaints** | Intensity or seriousness of reported issues |
| **20.** | **RVIS** | Eye health assessment metric or index |
| **21.** | **Ocularsymptomsobservedlately** | Recent observations of eye-related symptoms |
| **22.** | **Symptomsobservingatleasthalfofthetimes** | Frequency of experiencing symptoms |
| **23.** | **Complaintsfrequency** | Rate of complaint occurrence |
| **24.** | **frequencyofdryeyes** | Occurrence rate of dry eye symptoms |
| **25.** | **Schimers1Lefteye** | Measurement of tear production in the left eye using Schirmer's test |
| **26.** | **Schimers1righteye** | Measurement of tear production in the right eye using Schirmer's test |
| **27.** | **Schimers2Lefteye** | Measurement of tear production after anesthesia in the left eye using Schirmer's test |
| **28.** | **Schimers2righteye** | Measurement of tear production after anesthesia in the right eye using Schirmer's test |

**5.2 Exploratory Data Analysis**

Exploratory Data Analysis (EDA) serves as a crucial initial phase in the machine learning workflow, aiming to comprehensively examine and comprehend the data's properties before engaging in modeling endeavors. Through EDA, analysts seek to elucidate fundamental data characteristics, including distribution, inter-variable correlations, and discernible patterns or irregularities. This preliminary exploration is pivotal as it furnishes a profound insight into the dataset, facilitating informed decisions regarding feature manipulation, data preprocessing, and model selection.

By undertaking EDA, researchers can pinpoint and rectify any missing or erroneous data, outliers, or inconsistencies, thus ensuring a more robust foundation for subsequent machine learning tasks.

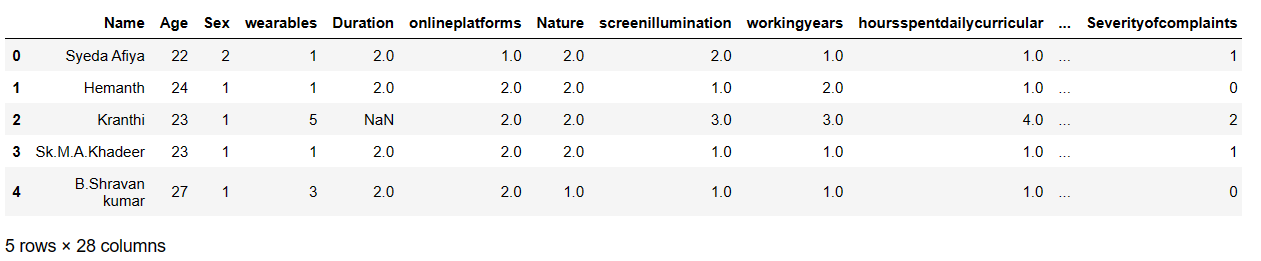
* **Information about the Features & their data types**

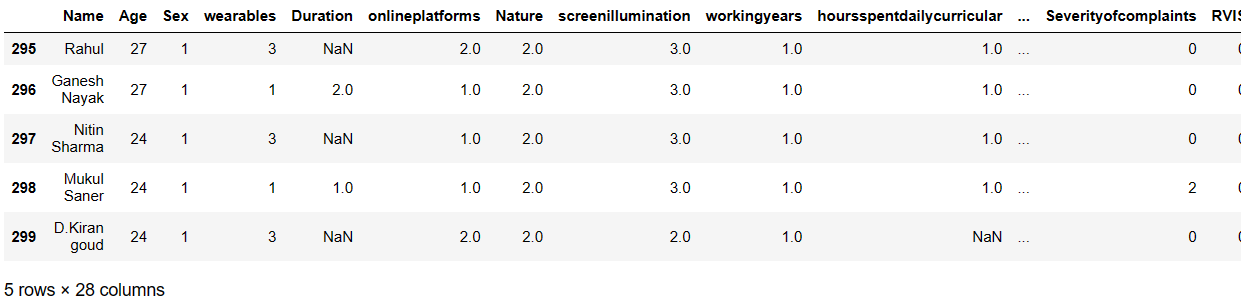


**Observations:**

* Figure 1 outlines the data types and non-null entry counts of the dataset columns. Upon initial review, it's evident that the 'Name' column is categorical, while others appear numerical but are also categorical. Considering the survey form, we infer that these numerical values represent coded options. Further analysis of unique values will validate this observation.
* In the given dataset, there are missing values. Addressing these missing values requires data pre-processing, where we either remove or replace them.
* Within the dataset, multiple target variables include “Schimers1Lefteye, Schimers1righteye”, and “Schimers2Lefteye”, “Schimers2righteye”. These variables likely represent different measurements or aspects related to eye examinations

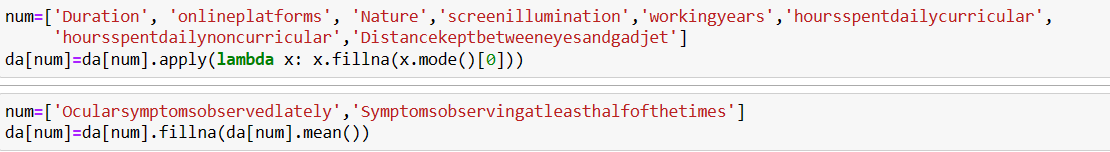
**5.2.1 Slicing and Dicing**





**Observation:**

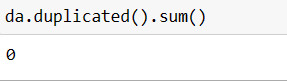
The presence of numerous null values in the dataset indicating there are outliers necessitates data cleaning, requiring the replacement of these null values with appropriate statistical measures such as mean, median, or mode.



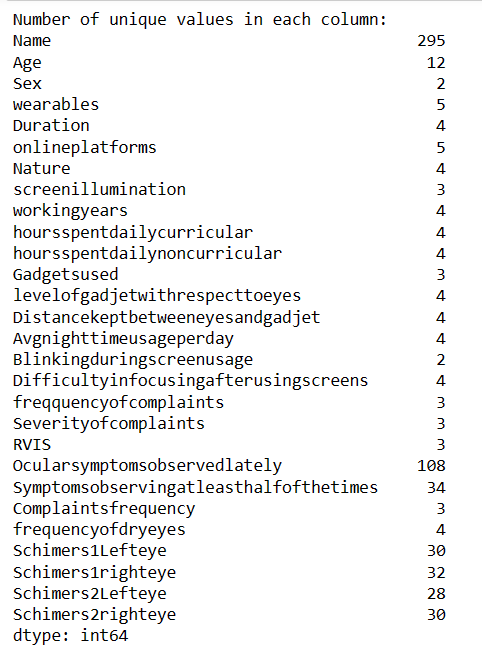
* 'Duration','onlineplatforms','Nature','screenillumination','workingyears','hoursspentdailycurricular','hoursspentdailynoncurricular','Distancekeptbetweeneyesandgadjet' , columns are replaced by mode.
* ‘Ocularsymptomsobservedlately’,’Symptomsobseringatleasthalfofthetimes are replaced by mean.

**5.2.2 Checking for Data Consistency**

* **No duplicates found.**

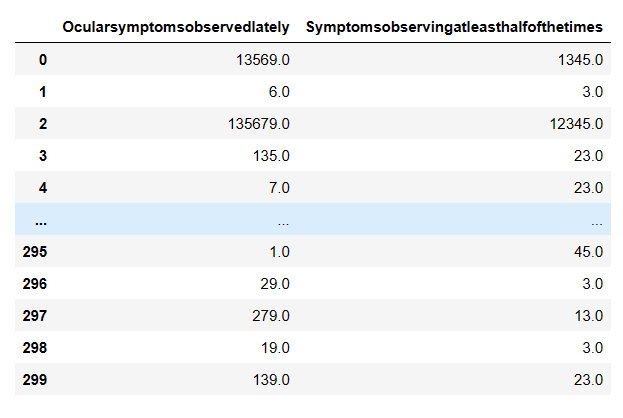


• **Unique Values**



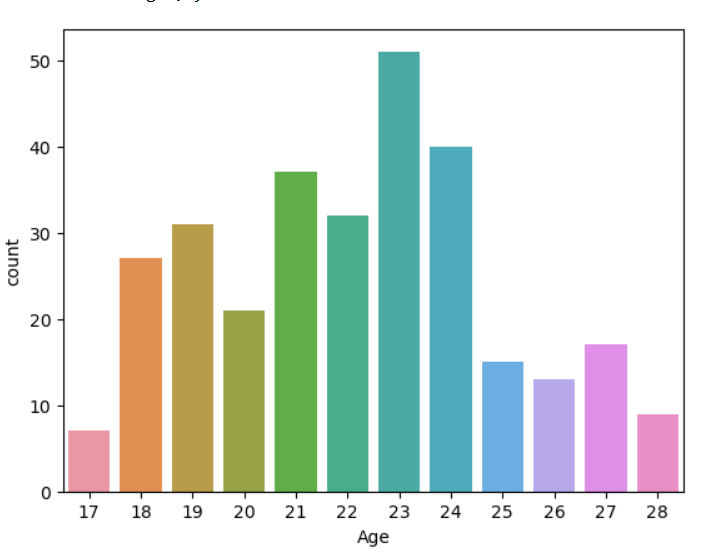
**Observations:**

* From the analysis of unique value counts (Fig), it is evident that the 'Name' column has nearly all unique values. Given that it serves as an identifier, it can be safely dropped from the dataset as it does not contribute to the analysis of eye health.
* Excluding the target columns ('Schimers1Lefteye', 'Schimers1righteye', 'Schimers2Lefteye', 'Schimers2righteye') and the mentioned columns ('Ocularsymptomsobservedlately' , 'Symptomsobservingatleasthalfofthetimes'), the remaining columns exhibit a limited number of unique values. This suggests that these columns likely represent categorical variables that have been label encoded. Each unique numerical value corresponds to a specific option provided in the survey form.
* Regarding 'Ocularsymptomsobservedlately' and 'Symptomsobservingatleasthalfofthetimes', these columns likely capture information about ocular symptoms reported by participants. The presence of a variety of symptoms and their frequency of occurrence may provide valuable insights into the prevalence and severity of eye-related issues among the surveyed individuals. Further analysis of these columns could reveal patterns and associations between reported symptoms and other variables in the dataset.



**5.2.3 Data Visualization**

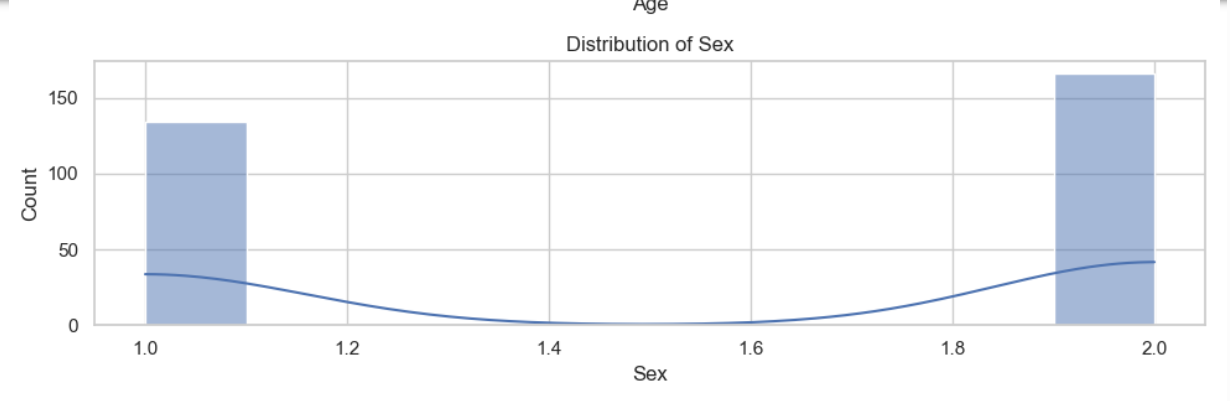
**Distribution of Age:**



**Observation**:

* The graph displays a count of individuals across different age groups, ranging from 17 to 28 years old.
* The provided histogram illustrates that individuals aged 23 constitute the largest proportion among the dataset's population.
* Age groups 21 and 24 follow closely behind, with slightly lower counts than the 23-year-olds.
* Age groups of 17, 26, and 28 have significantly lower counts compared to the other ages, suggesting fewer individuals fall into these age brackets within the dataset.

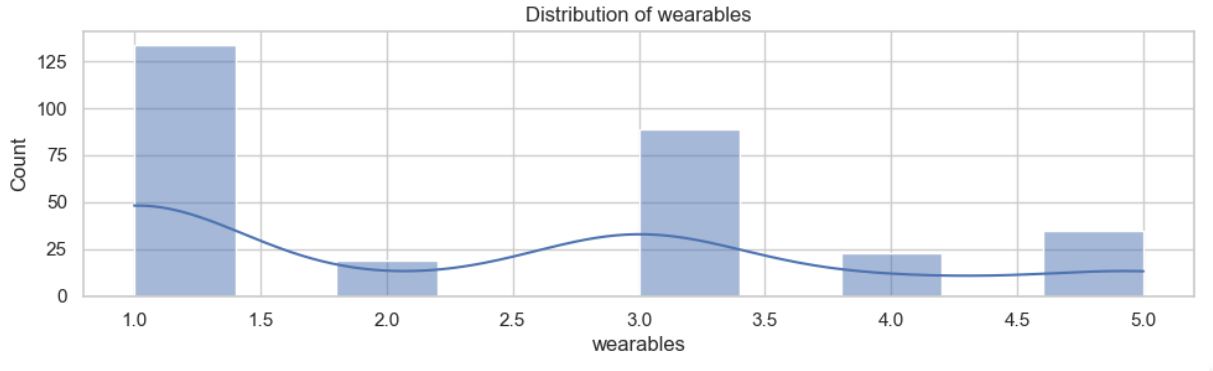
**Distribution of Sex**:



**Observation:**

1. The graph exhibits a **bimodal distribution**, with peaks at approximately **1.0** (men) and **2.0** (women).
2. The graph says that more number of women count in the above data

**Distribution of wearables:**

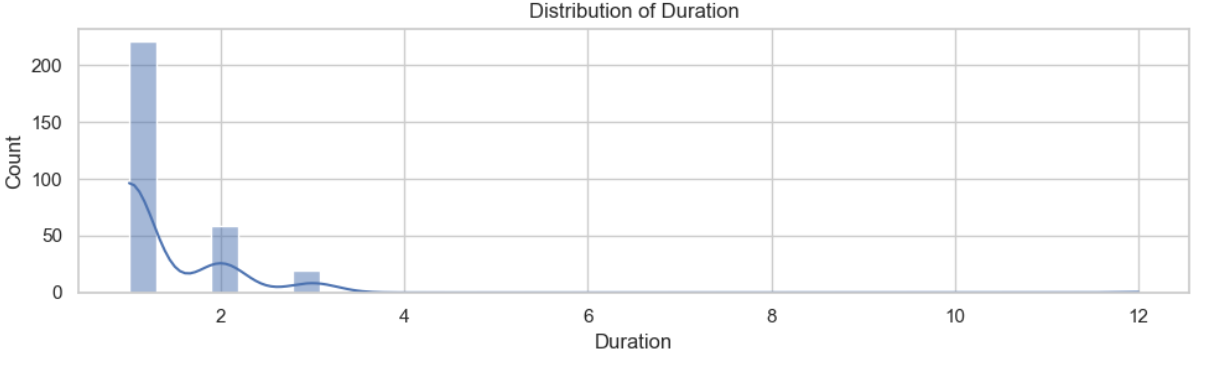


**Observation:**

Wearables have many subcategories,

* 1-Glasses only: : This category has the highest count, nearly 140, indicating that it’s the most common type of wearable in the dataset. It suggests that Type 1.0 might be a standard or basic model that is widely used.
* 2-Contact lens
* 3-None: **With a count close to 80, this type is the second most prevalent in the dataset. This implies that there are close to 80 participants who do not wear any glasses or contacts.**
* 4-Both,at one or other
* 5-Not aplicable
* The histogram above indicates a significant prevalence of individuals wearing glasses compared to other subcategories, notably with "None" being the next prominent category, indicating individuals not wearing any wearables.

**Distribution of Duration:**



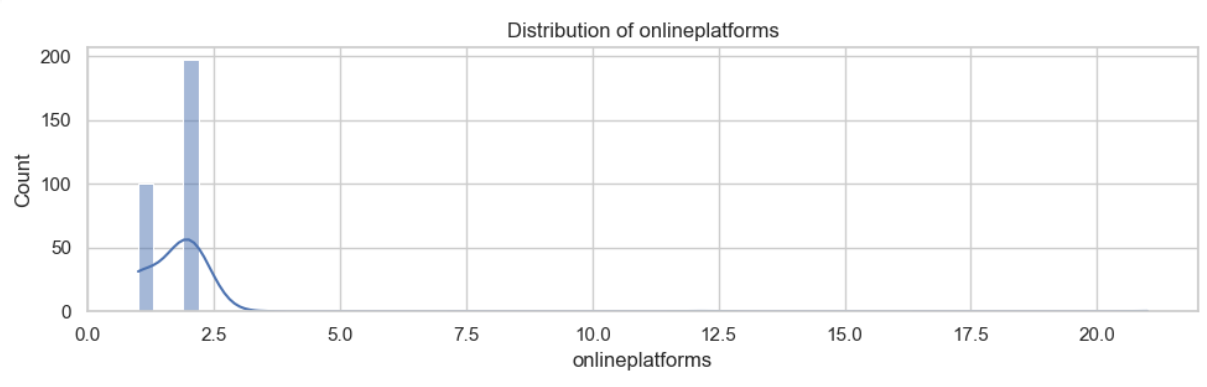
**Observation:**

In Duration,

1. less than 5 years
2. 5 to 10 years
3. Greater than 10 years

The histogram above indicates a significant prevalence of individuals duration of usage of prescription glasses or lenses is less than 5 years compared to other subcategories, notably with 5-10 years being the next one.

**Distribution of online platforms:**

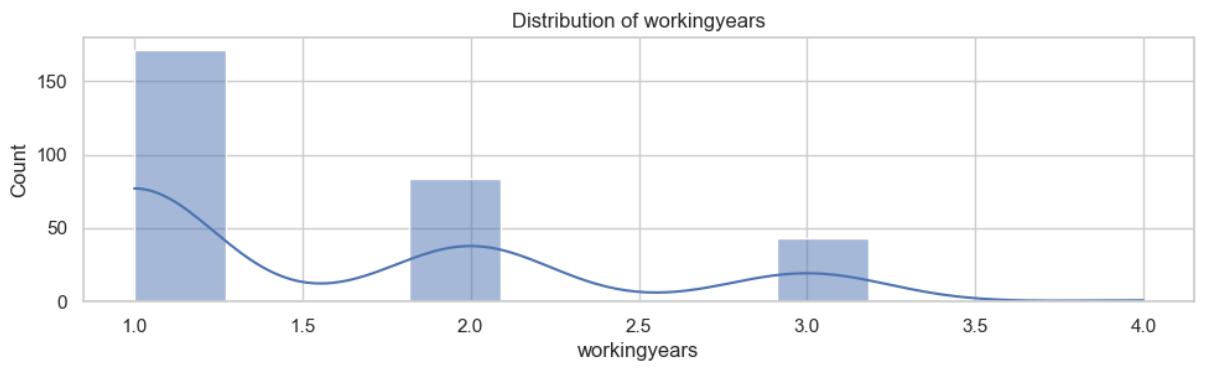


**Observation:**

It indicates whether a person is using any online platforms to attend any webinars/classes,

1. YES
2. NO

The histogram indicates that the majority of people are not utilizing the platforms for any classes or webinars.

**Distribution of working years:** 

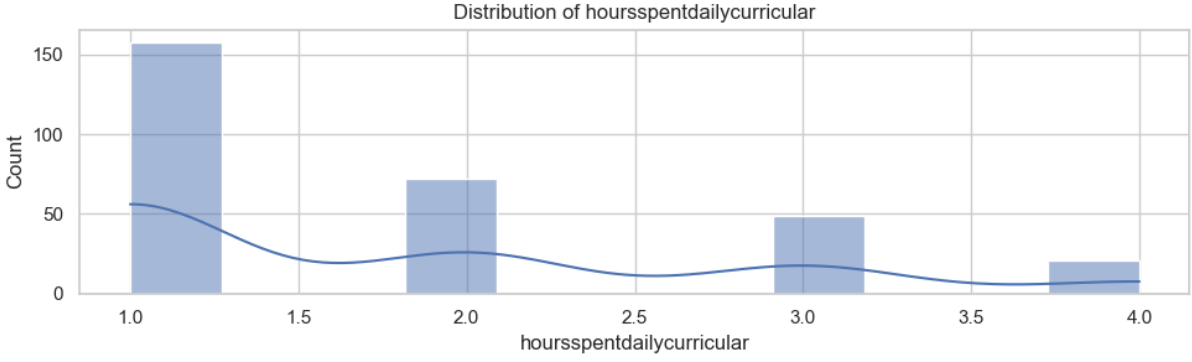
**Observation:**

This depicts for how many years a person have been working with computer on daily basis,

1. less than 5 years
2. 5 to 20 years
3. Greater than 10 years

The above histplot represents most of the people are working for less than 5 years and 5-10 years is the next prominent one.

**Distribution of hours spent daily curricular:**



**Observation:**

This represents how much out of the time total is spent for online classes in a day,

1-2 to 4 hours

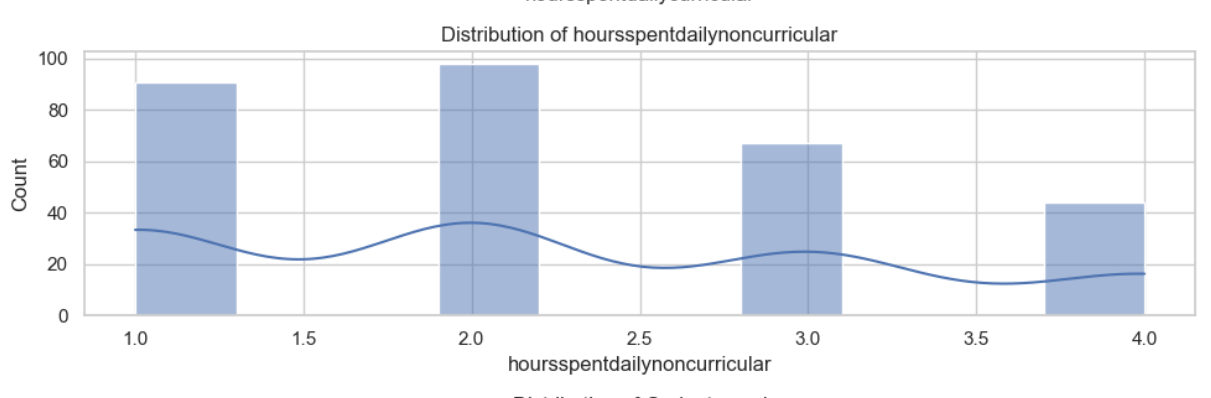
2-4 to 6 hours

3-6 to 8 hours

4-8 to 10 hours

As we can observe there is a notable concentration observed in the 2 to 4 hours category, indicating that most individuals spend this duration on online classes.

**Distribution of hours spent daily non-curricular:**



**Observation:**

This represents how much of screen time for reason other than online classes in a day,

1-less than 2 years

2-2 to 4 hours

3-4 to 6 hours

4-6 to 8 hours

As we can observe there is a notable concentration observed in the 2 to 4 hours category, indicating that most individuals spend this duration on activities other than online classes.

**ScreenIllumination**

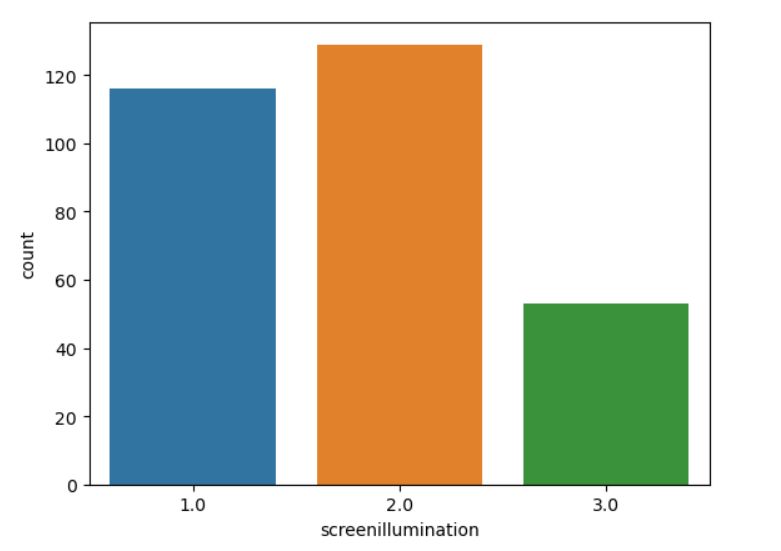
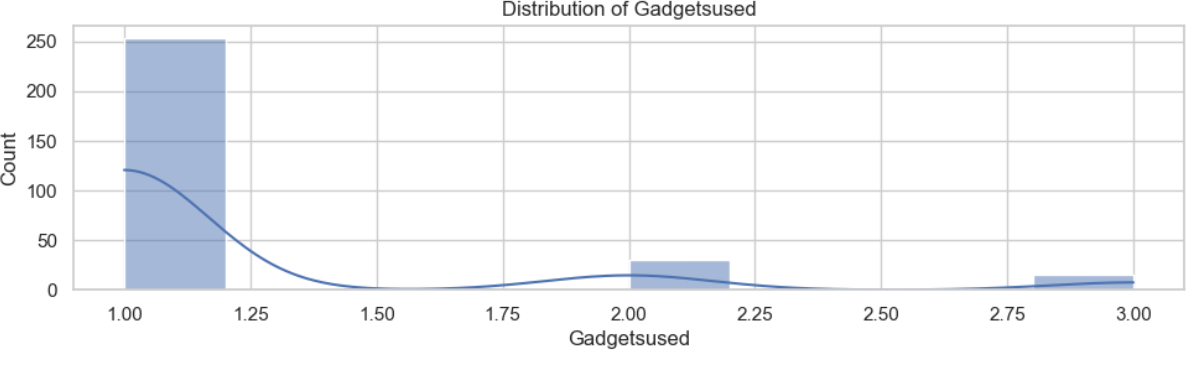
****

Fig : Graph representing the value count of screenillumination column.

Observation:

The analysis of screen illumination levels reveals notable variations in their respective counts. At the lowest level of illumination, categorized as 1.0 (Less than 25%), the count is just under 120, indicating a relatively high occurrence. Conversely, for level 2.0 (25 to 30%), the count exceeds 120, suggesting that this level of illumination results in the highest count among the three categories. In contrast, at level 3.0 (greater than 50%), the count is approximately 40, significantly lower than the other two levels. This implies that higher levels of screen illumination correlate with lower counts of the variable, indicating a potential trend in user preferences or behavior regarding screen brightness.

**Distribution of Gadget used**:



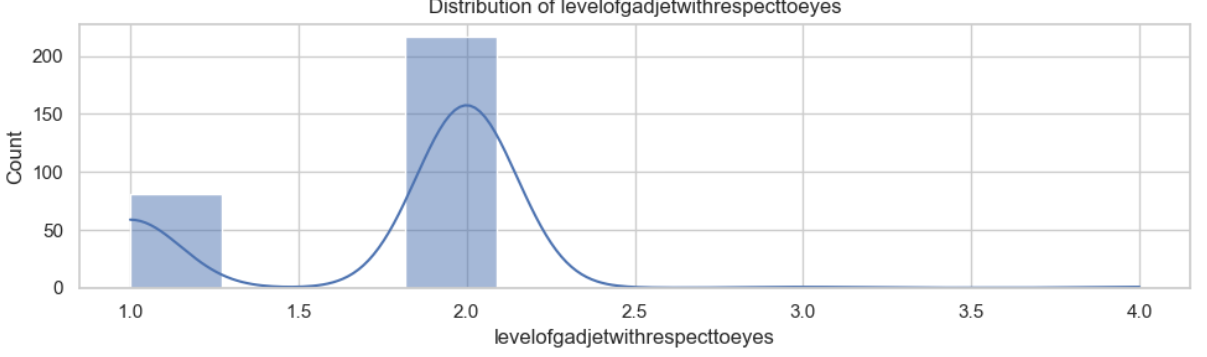
**Observation:**

This indicates what gadget is used most of the times,

1. Mobiles
2. Laptop
3. Tablet

As we can observe from above histplot that most of the people prefer using mobiles most of the time.

**Distribution of level of gadget with respect to eyes:**



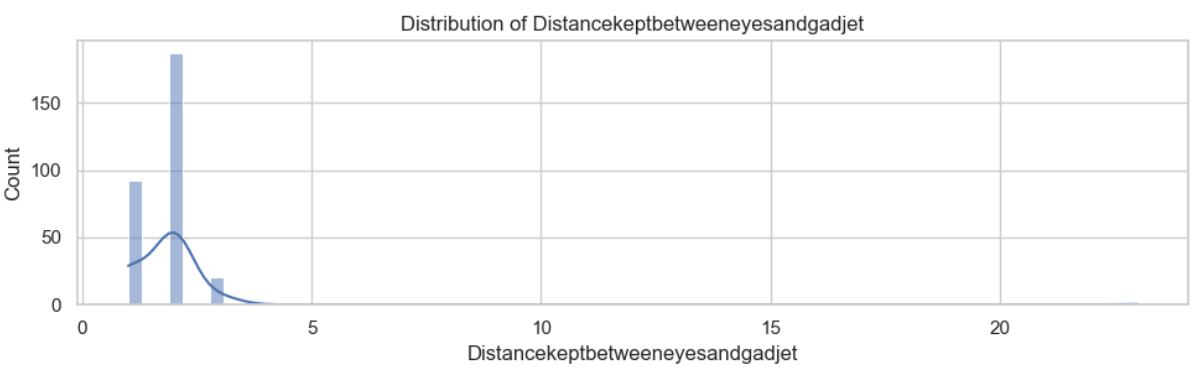
**Observation:**

The histogram suggests the position of gadgets relative to the eyes, with options including

1. Above the level of eyes
2. At the level of eyes

It indicates that the majority of individuals are using gadgets at the level of eyes, potentially contributing to eye-related issues.

**Distribution of Distance Kept between eye and gadget:**



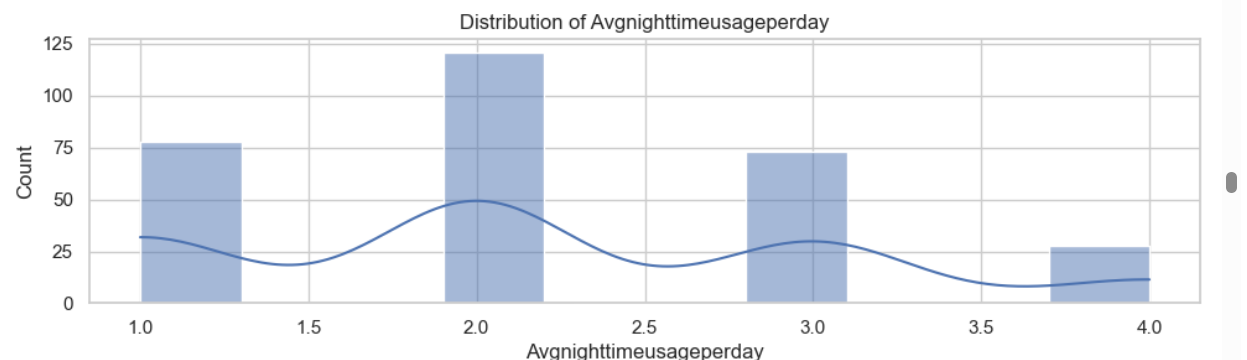
**Observation:**

It indicates the distance between the eyes and gadget more than half od the times,

1. less than 25cm
2. 25 to 40 cm
3. More than 40 cm

As we can observe most of the people are keeping the distance between 25 to 40 cm between their eyes and gadgets.

**Distribution of Average night time usage per day:**



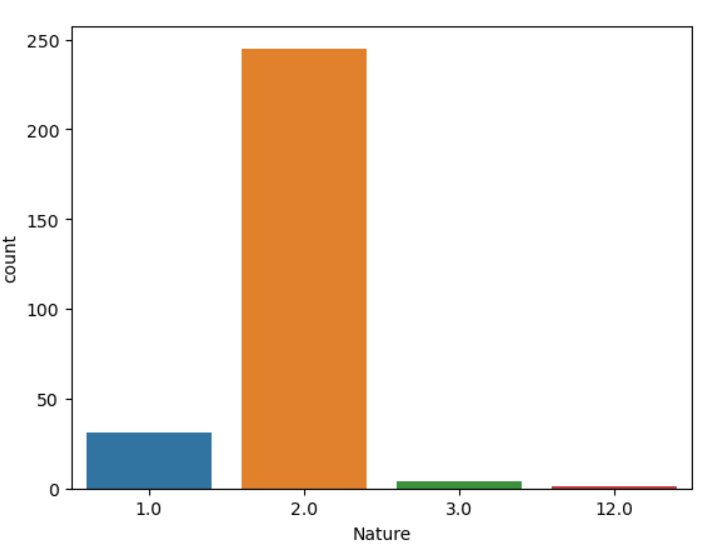
**Observation:**

It indicates the average night time usage of the gadget per day,

1. less than 1 hour
2. 1 to 2 hours
3. 2 to 4 hours
4. More than 4 hours

As we can observe 1 to 2 hours is the most significant usage of the gadgets at night.

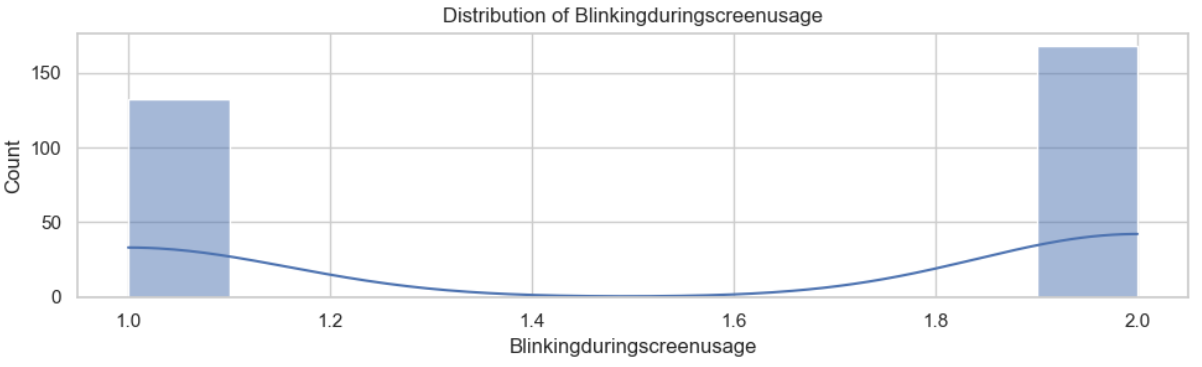
**Nature**

****

**Observation:**

* The frequency distribution analysis highlights that the majority of respondents report interrupted usage of their computers, as evidenced by a higher count in the "2.0 - Interrupted usage" category compared to "1.0 - Continuous usage."
* The prevalence of interrupted usage over continuous usage is a notable trend among the surveyed users, indicating a common pattern in their computer usage habits.
* The presence of values other than 1 and 2 in the frequency distribution graph suggests potential errors in the dataset. Given that the survey form only provides two options for computer usage, these outliers may indicate inaccuracies or inconsistencies in the data collection process.
* Considering the low frequency of these outliers, it is advisable to remove them from the dataset during the data cleaning process. This will help improve the accuracy and reliability of the data for subsequent analysis and modeling

**Distribution of Blinking during screen usage:**



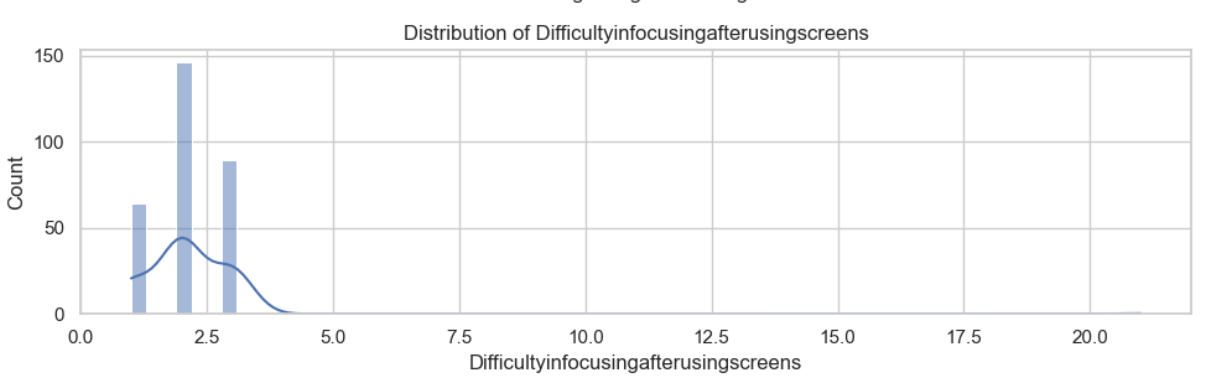
**Observation:**

It indicates whether a person keep remainding themselves to blink during digital screen usage,

1. YES
2. NO

From above histplot we can say that most of the people ignore blinking of their eyes during the usage of screen.

**Distribution if Difficulty info using after using screens:**



**Observation:**

It indicates after using screen the difficulty in focusing from one distance to another more than half of the times,

1. Yes
2. NO
3. MAY BE

The histplot represents most of the people does not feel difficulty in focusing from one place to another after usage of digital screen.

**Distribution of frequency of dry eyes:**



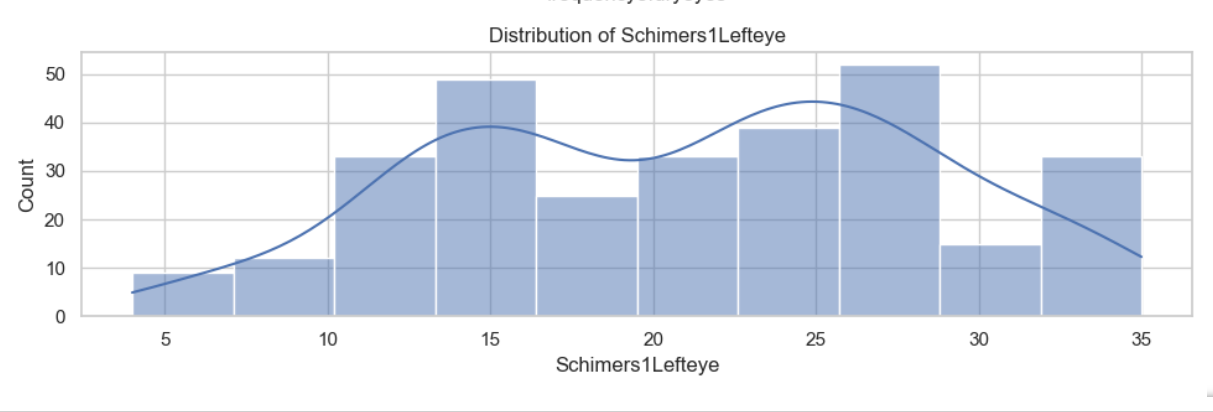
**Observation:**

It indicates the frequency of experiencing the dry eyes,

1. Often
2. Sometimes
3. Never

As we can observe Sometimes is significantly high.

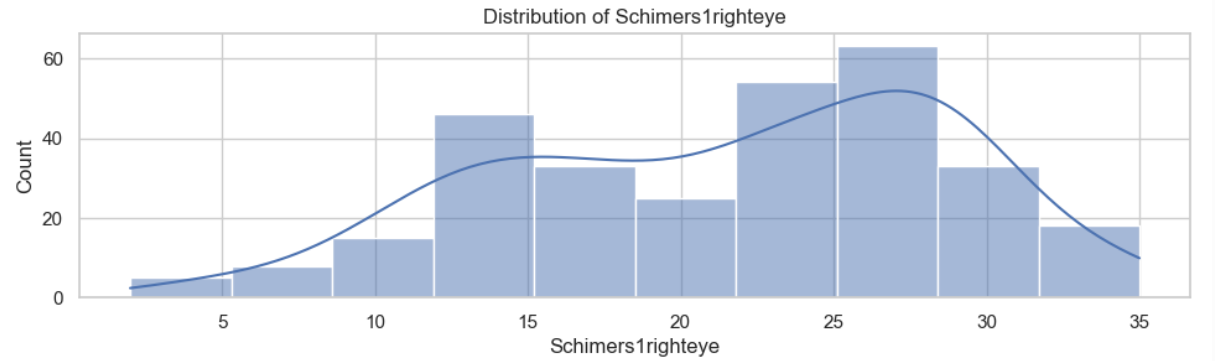
**Distribution of Schimers1lefteye:**



**Observation:**

It's important to analyze and address the data trends regarding Schimers1lefteye levels among different age groups. The observation highlights a notable prevalence of higher Schimers1lefteye levels among individuals aged between 26 and 29.

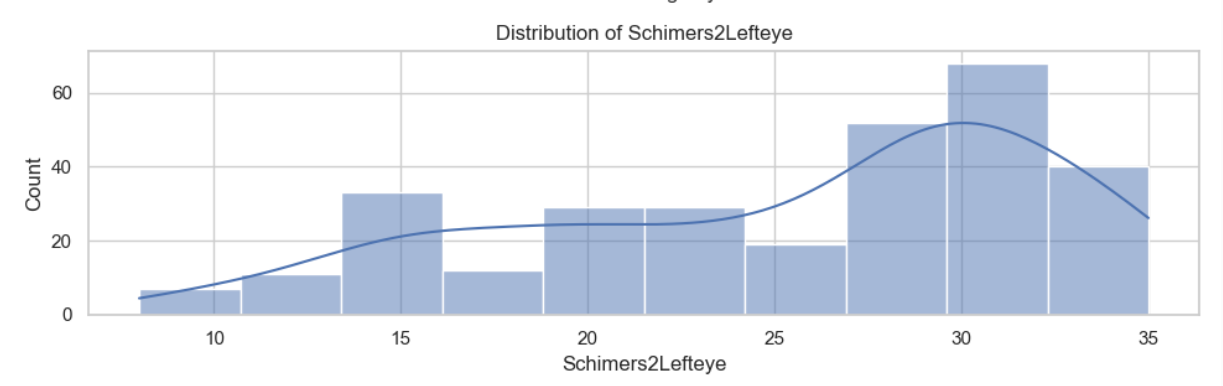
**Distribution of Schimers1righteye:**



**Observation:**

It's important to analyze and address the data trends regarding Schimers1righteye levels among different age groups. The observation highlights a notable prevalence of higher Schimers1lefteye levels among individuals aged between 25 and 28.

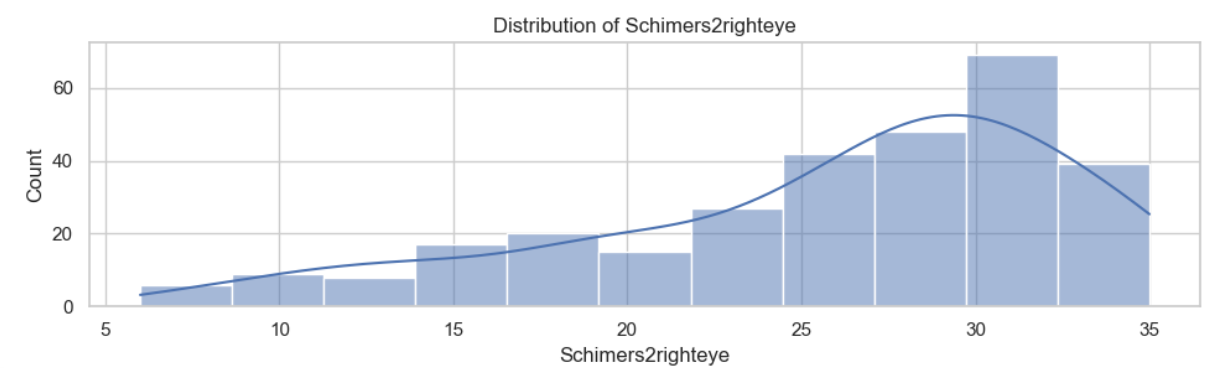
**Distribution of Schimers2lefteye:**



**Observation:**

* It's important to analyze and address the data trends regarding Schimers2lefteye levels among different age groups.
* The observation highlights a notable prevalence of higher Schimers1lefteye levels among individuals aged between 28 and 32.
* The bars initially increase in height, peak around the value of 30, and then decrease.

**Distribution of Schimers2righteye:**



**Observation:**

It's important to analyze and address the data trends regarding Schimers1lefteye levels among different age groups. The observation highlights a notable prevalence of higher Schimers1lefteye levels among individuals aged between 29 and 32.

**Dividing Multiple Target Variables**

Recognizing the complexities involved in predicting multiple target variables simultaneously, our approach involves developing four distinct models, each tailored to predict one of the four target variables. These models are designed to utilize the same set of input variables, ensuring consistency and comparability across predictions. This strategy allows for a focused and specialized approach to address the unique characteristics and dependencies present in each target variable, ultimately enhancing the predictive accuracy and interpretability of our analyses.

### 6.ALGORITHMS

1. **Random Forest Classifier in Python**
2. **KNeighborsClassifier**

**3.Decision Tree Regression**

### 7.Implementation

# 7.1 KNeighborsClassifier

# The data set is divided into training and testing sets using the train\_test\_split 80% of the data is used for the training (240 samples) , and 20 % is reserved for testing (60 samples)

# The K Nearest Neighbors (KNN) algorithm is chosen for classification, with 11 neighbors, Euclidean distance metric, and p=2 (indicating the use of Euclidean distance).

# F1\_score:

# 

# The macro-average F1 score is approximately 0.28, indicating moderate performance.

# Prediction results:

# array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 2., 1.,

# 3., 1., 1., 1., 2., 2., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,

# 1., 1., 1., 4., 1., 1., 1., 1., 1., 1., 3., 3., 1., 1., 1., 1., 1.,

# 3., 1., 2., 1., 1., 1., 1., 1., 3.])

# Accuracy percentage: 43.333333333333336

# The KNN classifier demonstrates moderate performance in predicting the Schirmer test results based on the given features. While the accuracy is above chance level, there is room for improvement, as indicated by the confusion matrix and F1 score. Further optimization of the model parameters or exploring alternative algorithms may enhance the predictive capability

**7.2 Random Forest Classifier**

**0.9649805447470817**

**Classification Report**

**precision recall f1-score support**

**0 0.96 0.98 0.97 132**

**1 0.98 0.95 0.96 125**

**accuracy 0.96 257**

**macro avg 0.97 0.96 0.96 257**

**weighted avg 0.97 0.96 0.96 257**

**Accuracy: 96.5%**

**Observations**

* The Random Forest classifier achieves an accuracy of approximately 96.5%, indicating that the model correctly predicts the target variable (heart condition) for most of the instances in the test set.
* The Random Forest classifier demonstrates strong performance in predicting heart conditions based on the given features.
* The high accuracy, precision, recall, and F1-scores indicate that the model is robust and reliable.

**7.3 Decision Tree Classifier**

**DecisionTreeClassifier**

**DecisionTreeClassifier(max\_depth=7, min\_samples\_split=10, random\_state=42)**

**precision recall f1-score support**

**4.0 0.00 0.00 0.00 0**

**6.0 0.00 0.00 0.00 1**

**7.0 0.00 0.00 0.00 1**

**8.0 0.50 0.50 0.50 2**

**10.0 0.00 0.00 0.00 0**

**11.0 0.00 0.00 0.00 4**

**12.0 0.50 0.29 0.36 7**

**13.0 0.00 0.00 0.00 2**

**14.0 0.14 0.20 0.17 5**

**15.0 0.00 0.00 0.00 0**

**16.0 0.00 0.00 0.00 4**

**17.0 0.00 0.00 0.00 0**

**18.0 0.20 0.14 0.17 7**

**20.0 0.00 0.00 0.00 0**

**21.0 0.00 0.00 0.00 0**

**22.0 0.50 0.31 0.38 13**

**23.0 0.00 0.00 0.00 0**

**24.0 0.00 0.00 0.00 1**

**25.0 0.00 0.00 0.00 0**

**26.0 0.25 0.17 0.20 6**

**27.0 0.00 0.00 0.00 0**

**28.0 0.33 0.08 0.12 13**

**30.0 0.00 0.00 0.00 3**

**...**

**accuracy 0.17 75**

**macro avg 0.11 0.08 0.09 75**

**weighted avg 0.28 0.17 0.21 75**

**Accuracy :17%**

**Observations:**

* The overall accuracy of the classifier is 17%, which is very low. This means that the model correctly predicts the class only 17% of the time.
* The model's performance is poor across all metrics, indicating that it is unable to effectively classify the data into the correct classes.
* There might be several reasons for this poor performance, such as inadequate feature selection, insufficient data, or improper parameter tuning.

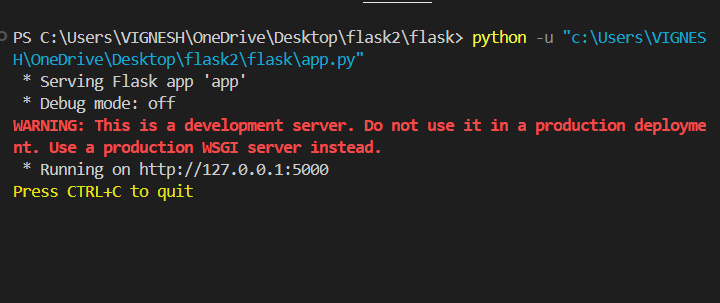
**Overall Observation:**

* We can clearly see **Random Forest Classifier** has outperformed all other models.
* Accuracy is excellent for Random Forest Classifer.
* So the below model is built based on this classifier.

**8. Final implementation of model Using Flask Website**

**RandomForestClassifier**

**Step 1: run the python code app.py**

****

**Create Flask App:**

* Import the Flask class and instantiate it to create a Flask app instance.
* Define routes using the @app.route('/myurl') decorator to specify URL endpoints for different functionalities.

**Write Random Forest Classifier and Model. pkl File:**

* Create a Python script named random\_forest\_classifier.py.
* Write Python code in this script to implement the random forest classifier for predicting the Schirmer test results.
* Once the classifier is trained and ready, save it into a file named model.pkl using the pickle library.

**Implement Random Forest Classifier**:

* In the random\_forest\_classifier.py script, write code to load the dataset, preprocess it, train the random forest classifier, and evaluate its performance.
* After training, save the trained classifier to the model.pkl file using the pickle.dump() function.

**Create Index.html File:**

* Create an HTML file named index.html to serve as the user interface for inputting data.
* Design the HTML page to contain input fields for all 23 attributes required for the Schirmer test prediction.

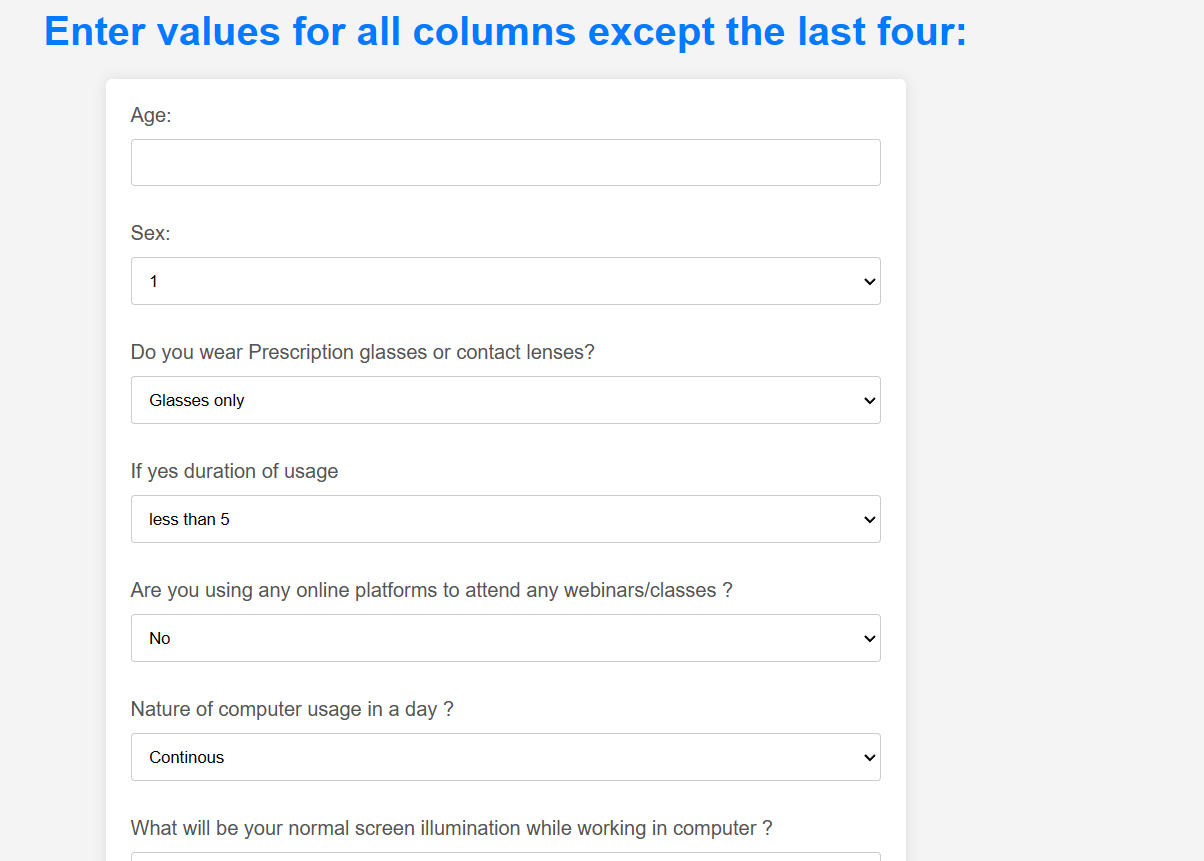
**Implement Flask Routes:**

* Define Flask routes to handle requests from the client-side.
* For example, create a route to render the index.html template and another route to handle form submissions.

**Run Flask App:**

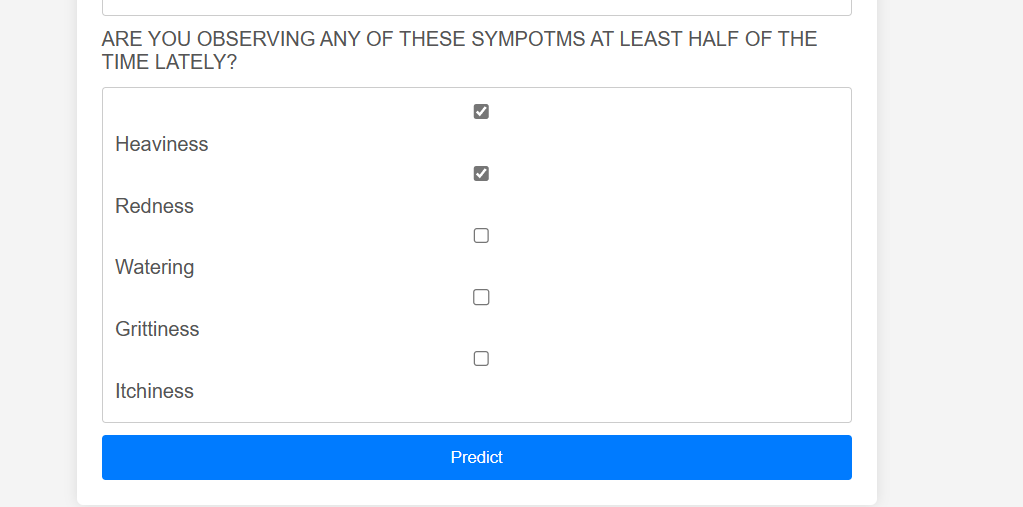
* Run the Flask app using the flask run command or by executing the Python script containing the Flask app.
* Access the Flask app through a web browser to interact with the user interface and make predictions.

**Step 2: open the link in browser:**

****

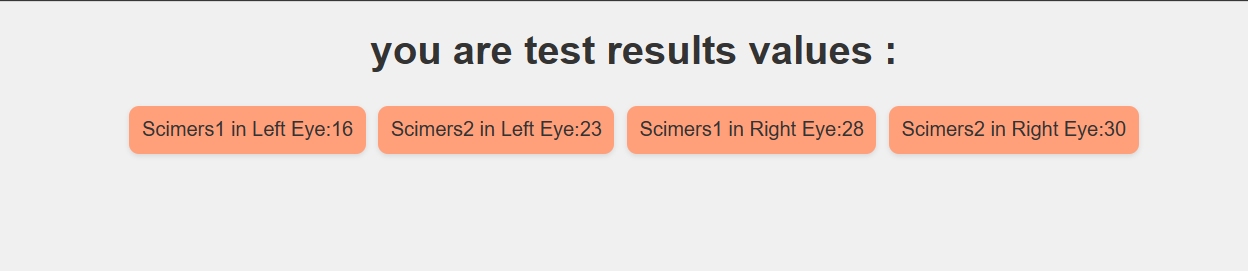
* Once the web application is open, you'll see input fields corresponding to each attribute required for predicting Schirmer test values.
* Enter the values for each attribute in the respective input tabs.
* Additionally, make selections from any choice boxes provided, if applicable, based on the options available.

**Step three click on prediction button**

****

After entering all the necessary input values, click on the "Prediction" button. This action triggers the Flask app to process the input data using the trained model

**Step 4: get the predicted values of schrimerslefteye and schimersrighteye**

****

Upon clicking the "Prediction" button, the Flask app will utilize the trained random forest classifier to predict the Schirmer test values for both the left eye and the right eye.

The predicted values will be displayed on the web page.

**Final Observation**

* Upon receiving the predicted Schirmer test values
* It becomes evident that the predicted values for the Schirmer test of the right eye (Schirmer1 right eye and Schirmer2 right eye) tend to be higher compared to those of the left eye.
* This observation suggests potential differences in tear production between the two eyes

**Conclusion:**

In our study, we employed several popular machine learning algorithms, including Random Forest Classifier, Decision Tree, and K-Nearest Neighbors Classifier. Upon evaluation, we found that the Random Forest Classifier exhibited higher accuracy levels than the other algorithms. Consequently, we chose the Random Forest Classifier for the final implementation of our model, which yielded the digital eye strain results. This decision was driven by the algorithm's demonstrated ability to provide reliable and accurate predictions, making it well-suited for our specific task.