Visvesvaraya Technological University

Belagavi, Karnataka-590 014

A Internship Report on

"MOBILE PRICE CLASSIFICATION"

Internship Report submitted in partial fulfillment of the requirement for the
award of the degree of

Master of Computer Applications

Submitted by

CHETHAN G
(1AY22MC023)

Under the Guidance of
Prof. PALLAVI M O

Acharya Institute of Technology
Acharya Dr. Sarvepalli Radhakrishnan Road, Soladevanahalli,
Bangalore-560107, India

2023-2024

Master of Computer Applications,

Acharya Institute of Technology

CERTIFICATE

This is to certify that Mr. **CHETHAN G** bearing USN **1AY22MC023** has satisfactorily completed his/her third semester Internship work entitled **"MOBILE PRICE CLASSIFICATION"** as a partial fulfillment for the award of a Master of Computer Applications degree, during the year 2023-2024 under joint supervision.

Internal Guide Head of the Department Principal

Examiners signature with Date:

1. External Examiner

2. Internal Examiner

INTERNSHIP CERTIFICATE

edunet
foundation

CERTIFICATE

OF COMPLETION

awarded to

Chethan G

for successfully completing 4-week Internship, leveraging SkillsBuild & IBM Cloud Platform in

Emerging Technologies (AI & Cloud)

from 6th Nov 2023 to 4th Dec 2023

This program was conducted by **Edunet Foundation**

Nagesh Singh

Executive Director
Edunet Foundation

College Name: Acharya Institute of Technology
Email ID: chethang.22.mcav@acharya.ac.in
Students ID: 1AY22MC023

DECLARATION

I **Chethan G,** student of the 3rd semester of Master of Computer Applications, Acharya Institute of Technology, Bangalore-560107, bearing USN **1AY22MC023** hereby declare that the Internship project entitled **"MOBILE PRICE CLASSIFICATION"** been carried out by me under the supervision of Internal Guide **Prof. PALLAVI M O** submitted in partial fulfillment of the requirements for the award of the degree of Master of Computer Applications by the Visvesvaraya Technological University during the academic year 2023-2024. This report has not been submitted to any other organization/university for any award of degree or certificate.

Name: **Chethan G**

Sign:

ACKNOWLEDGEMENT

It is a great pleasure to acknowledge all the people who have contributed towards completing my Internship, who all are helping me to do this Internship directly or indirectly for their support and guidance.

I am very much thankful to the principal **Dr. Rajath Hegde** for providing academic support for the success of the Internship.

I am very grateful to **Dr. Ratnakirti Rao** our respected Head of the Department, dept of MCA, Acharya Institute of Technology, Bangalore for providing the academic environment and their support for the Internship and keeping me on the correct path.

My heartful gratitude to **IBM** for their constant support, and for allowing me to carry out my internship work.

My special thanks to my internal guide **Prof. Pallavi M O**, dept of MCA, Acharya Institute of Technology, Bangalore, for her suggestions, cooperation, and constant support when it is needed for my Internship.

My special thanks to all the Professors of MCA and my sincere feelings to my parents and friends for giving me immense support and motivation and also who are supporting me directly or indirectly during my completion of the Internship.

.

CHETHAN G

ABSTRACT

Understanding the relationship between a mobile phone's features and its selling price is crucial in the competitive market. For new entrants lacking expertise in advanced machine learning, accurately estimating product prices poses a challenge. This study presents a simplified approach to price range estimation using sales data from various sources. The dataset includes diverse mobile phones with features like RAM, internal memory, camera specifications, display size, and processor type.

Instead of complex algorithms, basic statistical analysis and visualization techniques are employed to identify patterns and trends. By analyzing feature distribution across different price categories and exploring correlations, actionable insights for price range estimation are derived. This approach aims to empower new businesses with a pragmatic framework for making informed pricing decisions. By understanding the importance of different features in determining price ranges, businesses can strategically position their products in the market and compete effectively.

TABLE OF CONTENT

1. INTRODUCTION

In today's highly competitive business environment, understanding consumer behavior and staying abreast of evolving shopping trends are imperative for organizations seeking sustained growth and competitiveness. The rapid proliferation of data and advancements in analytics technologies offer unprecedented opportunities to gain deep insights into consumer preferences, purchasing patterns, and market dynamics. This project aims to harness the power of data analytics to conduct a thorough analysis of shopping trends, with the goal of providing actionable insights for businesses operating within a specific market segment or industry.

The significance of this analysis lies in its potential to uncover hidden patterns, correlations, and opportunities that traditional market research methods may overlook. By analyzing large volumes of transactional data, demographic information, and other relevant variables, we aim to identify key drivers influencing consumer behavior and purchasing decisions. Moreover, by examining historical trends and utilizing predictive modeling techniques, we seek to anticipate future shifts in consumer preferences, enabling organizations to proactively adapt their strategies and stay ahead of the competition.

This introduction sets the stage for a comprehensive exploration of shopping trends through data analytics. The subsequent sections will delve into the specific objectives, methodologies, findings, and implications of this analysis, ultimately providing valuable insights to guide strategic decision-making and drive business success in the dynamic retail landscape.

In addition to unveiling hidden patterns and correlations, this project aims to delve into the realm of sentiment analysis, leveraging natural language processing techniques to decode customer sentiments and perceptions embedded within reviews, feedback, and social media interactions. By tapping into the rich source of unstructured data, we can gain deeper insights into consumer sentiment towards products, brands, and shopping experiences, thereby enriching our understanding of customer preferences and sentiment-driven behaviours.

2. OBJECTIVE OF STUDY

The primary objective of this study is to develop a simplified approach to estimate the price range of mobile phones in the competitive market. Specifically, the study aims to:

1. Analyze the relationship between various features of mobile phones (such as RAM, internal memory, camera specifications, display size, battery capacity, and processor type) and their corresponding selling prices.
2. Utilize basic statistical analysis and visualization techniques, including histograms, scatter plots, and correlation matrices, to identify patterns and trends in the dataset.
3. Explore correlations between different features and price ranges to uncover actionable insights regarding the relative importance of each feature in determining price.
4. Investigate any potential nonlinear relationships or interactions between features that may affect price range estimation.
5. Evaluate the performance of the proposed simplified approach in estimating price ranges compared to more complex machine learning models.
6. Provide practical guidelines and recommendations for new entrants and businesses, such as Bob's mobile company, on pricing strategies and product positioning based on the insights gained from the analysis.
7. Offer insights into consumer preferences and market dynamics in the mobile phone industry, enabling businesses to adapt their product offerings and marketing strategies accordingly.
8. Contribute to the existing body of knowledge in pricing analytics and market research by demonstrating the effectiveness of simplified methodologies in addressing real-world challenges faced by businesses in competitive markets

3. METHODOLOGY

System requirements

Hardware Requirements:

- Processor (CPU): A multi-core processor (at least dual-core) for faster computation.
- RAM: At least 8 GB of RAM is recommended, but more is better for handling large datasets and complex analyses.
- Storage: Adequate storage space for datasets and any additional software you may install.

Software Requirements:

- Python
- Package Management: Use pip or conda to install Python packages. These tools help manage and install the libraries you'll use in your project.
- Jupyter Notebook: Install Jupyter Notebook to create and run your notebooks.
- Data Analysis Libraries: Install essential libraries like pandas, numpy, matplotlib, and seaborn for data analysis and visualization.

4. DETAILED EXPLANATION OF WORK DONE

Week 1: Artificial Intelligence and Machine Learning

During the first week of my internship, I delved into the fundamentals of Artificial Intelligence and Machine Learning (AIML). Under the guidance of experienced mentors, I grasped key concepts, including data preprocessing, model building, and evaluation techniques. Through hands-on projects and interactive sessions, I gained practical insights into AIML applications across various domains. Additionally, I completed a certification to solidify my understanding and skills in this field. This immersive experience provided a strong foundation for my journey into the world of AIML, equipping me with valuable knowledge and expertise.

Week 2: IBM Cloud

In the second week of my internship, I immersed in the realm of cloud computing. Through comprehensive training sessions, I learned about cloud infrastructure, deployment models, and services offered by leading providers. Hands-on exercises and practical assignments deepened my understanding of cloud concepts and their real-world applications. Under the guidance of seasoned professionals, I honed my skills in cloud deployment, management, and security practices. Additionally, I completed a certification to validate my proficiency in cloud technologies, enhancing my credentials for future endeavors. This week was instrumental in broadening my knowledge and expertise in leveraging cloud solutions effectively.

Week 3: Chat Bot and Data Analytics

During the third week of my internship, I continued to deepen my understanding of chatbots and data analytics basics, with a focus on utilizing IBM Watson Studio. Through tailored instruction and practical exercises, I refined my skills in developing advanced chatbot functionalities and applying data analytics techniques to extract actionable insights. Hands-on experience with IBM Watson Studio empowered me to manipulate and analyze data effectively, paving the way for informed decision-making. By completing a certification, I cemented my expertise in both chatbot development and data analytics, bolstering my credentials and readiness for tackling complex projects in the future.

Week 4: Project

During the fourth week of my internship, I embarked on a project centered around auto sales data analysis using Python. Leveraging my programming skills and knowledge acquired in previous weeks, I collected, processed, and analyzed data gathered from smartwatches. Through Python libraries and frameworks, I implemented algorithms to extract valuable insights related to users' activities, health metrics, and behavioral patterns. Visualizations aided in conveying findings effectively. This hands-on project honed my Python proficiency, data analysis techniques, and project management skills, providing me with practical experience in real-world data analysis applications, particularly in the burgeoning field of wearable technology.

5. SCREENSHOTS OF OUTCOME

➢ Importing the libraries

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

➢ Loading the Dataset

In [2]:

```
df_train = pd.read_csv("/kaggle/input/mobile-price-classification/train.csv")
df_test  = pd.read_csv("/kaggle/input/mobile-price-classification/test.csv")
df       = pd.concat([df_train,df_test])
df
```

Out[2]:

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_width	ram
0	842	0	2.2	0	1	0	7	0.6	188	2	...	756	2549
1	1021	1	0.5	1	0	1	53	0.7	136	3	...	1988	2631
2	563	1	0.5	1	2	1	41	0.9	145	5	...	1716	2603
3	615	1	2.5	0	0	0	10	0.8	131	6	...	1786	2769
4	1821	1	1.2	0	13	1	44	0.6	141	2	...	1212	1411
...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	1700	1	1.9	0	0	1	54	0.5	170	7	...	913	2121
996	609	0	1.8	1	0	0	13	0.9	186	4	...	1632	1933
997	1185	0	1.4	0	1	1	8	0.5	80	1	...	825	1223
998	1533	1	0.5	1	0	0	50	0.4	171	2	...	832	2509
999	1270	1	0.5	0	4	1	35	0.1	140	6	...	608	2828

Out[2]:

m_dep	mobile_wt	n_cores	...	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range	id
0.6	188	2	...	756	2549	9	7	19	0	0	1	1.0	NaN
0.7	136	3	...	1988	2631	17	3	7	1	1	0	2.0	NaN
0.9	145	5	...	1716	2603	11	2	9	1	1	0	2.0	NaN
0.8	131	6	...	1786	2769	16	8	11	1	0	0	2.0	NaN
0.6	141	2	...	1212	1411	8	2	15	1	1	0	1.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.5	170	7	...	913	2121	14	8	15	1	1	0	NaN	996.0
0.9	186	4	...	1632	1933	8	1	19	0	1	1	NaN	997.0
0.5	80	1	...	825	1223	5	0	14	1	0	0	NaN	998.0
0.4	171	2	...	832	2509	15	11	6	0	1	0	NaN	999.0
0.1	140	6	...	608	2828	9	2	3	1	0	1	NaN	1000.0

3000 rows × 22 columns

- Information about the Dataset

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 3000 entries, 0 to 999
Data columns (total 22 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   battery_power   3000 non-null   int64  
 1   blue            3000 non-null   int64  
 2   clock_speed     3000 non-null   float64
 3   dual_sim        3000 non-null   int64  
 4   fc              3000 non-null   int64  
 5   four_g          3000 non-null   int64  
 6   int_memory      3000 non-null   int64  
 7   m_dep           3000 non-null   float64
 8   mobile_wt       3000 non-null   int64  
 9   n_cores         3000 non-null   int64  
 10  pc              3000 non-null   int64  
 11  px_height       3000 non-null   int64  
 12  px_width        3000 non-null   int64  
 13  ram             3000 non-null   int64  
 14  sc_h            3000 non-null   int64  
 15  sc_w            3000 non-null   int64  
 16  talk_time       3000 non-null   int64  
 17  three_g         3000 non-null   int64  
 18  touch_screen    3000 non-null   int64  
 19  wifi            3000 non-null   int64  
 20  price_range     2000 non-null   float64
 21  id              1000 non-null   float64
dtypes: float64(4), int64(18)
memory usage: 539.1 KB
```

In [4]:
```
df.drop(['id'],axis=1,inplace=True)
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 3000 entries, 0 to 999
Data columns (total 21 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   battery_power   3000 non-null   int64  
 1   blue            3000 non-null   int64  
 2   clock_speed     3000 non-null   float64
 3   dual_sim        3000 non-null   int64  
 4   fc              3000 non-null   int64  
 5   four_g          3000 non-null   int64  
 6   int_memory      3000 non-null   int64  
 7   m_dep           3000 non-null   float64
 8   mobile_wt       3000 non-null   int64  
 9   n_cores         3000 non-null   int64  
 10  pc              3000 non-null   int64  
 11  px_height       3000 non-null   int64  
 12  px_width        3000 non-null   int64  
 13  ram             3000 non-null   int64  
 14  sc_h            3000 non-null   int64  
 15  sc_w            3000 non-null   int64  
 16  talk_time       3000 non-null   int64  
 17  three_g         3000 non-null   int64  
 18  touch_screen    3000 non-null   int64  
 19  wifi            3000 non-null   int64  
 20  price_range     2000 non-null   float64
dtypes: float64(3), int64(18)
memory usage: 515.6 KB
```

- Checking if there are any null values present in the dataset or not ?

```
In [7]: df.isnull().sum()

Out[7]: Customer ID                0
        Age                        0
        Gender                     0
        Item Purchased             0
        Category                   0
        Purchase Amount (USD)      0
        Location                   0
        Size                       0
        Color                      0
        Season                     0
        Review Rating              0
        Subscription Status        0
        Shipping Type              0
        Discount Applied           0
        Promo Code Used            0
        Previous Purchases         0
        Payment Method             0
        Frequency of Purchases     0
        dtype: int64
```

```
In [8]: df.duplicated().sum()

Out[8]: 0
```

- Drops all rows from the Data Frame where at least one element is missing and it modifies

```
In [6]: df.dropna(axis=0,inplace=True)
        df
```

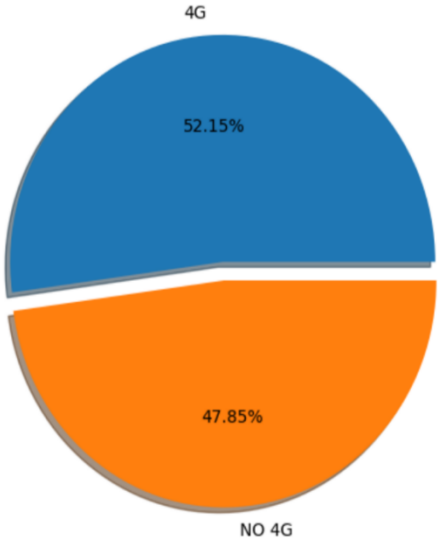
Out[6]:

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_
0	842	0	2.2	0	1	0	7	0.6	188	2	...	20	75
1	1021	1	0.5	1	0	1	53	0.7	136	3	...	905	19
2	563	1	0.5	1	2	1	41	0.9	145	5	...	1263	17
3	615	1	2.5	0	0	0	10	0.8	131	6	...	1216	17
4	1821	1	1.2	0	13	1	44	0.6	141	2	...	1208	12
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1995	794	1	0.5	1	0	1	2	0.8	106	6	...	1222	18
1996	1965	1	2.6	1	0	0	39	0.2	187	4	...	915	19
1997	1911	0	0.9	1	1	1	36	0.7	108	8	...	868	16
1998	1512	0	0.9	0	4	1	46	0.1	145	5	...	336	67
1999	510	1	2.0	1	5	1	45	0.9	168	6	...	483	75

2000 rows × 21 columns

- Distribution of two categories, 4G and NO 4G

In [9]:
```
labels = ['4G','NO 4G']
plt.pie(list_of_4G,labels=labels,shadow=True,autopct='%1.2f%%',radius=1.22,explode=[.1,0])
plt.show()
```



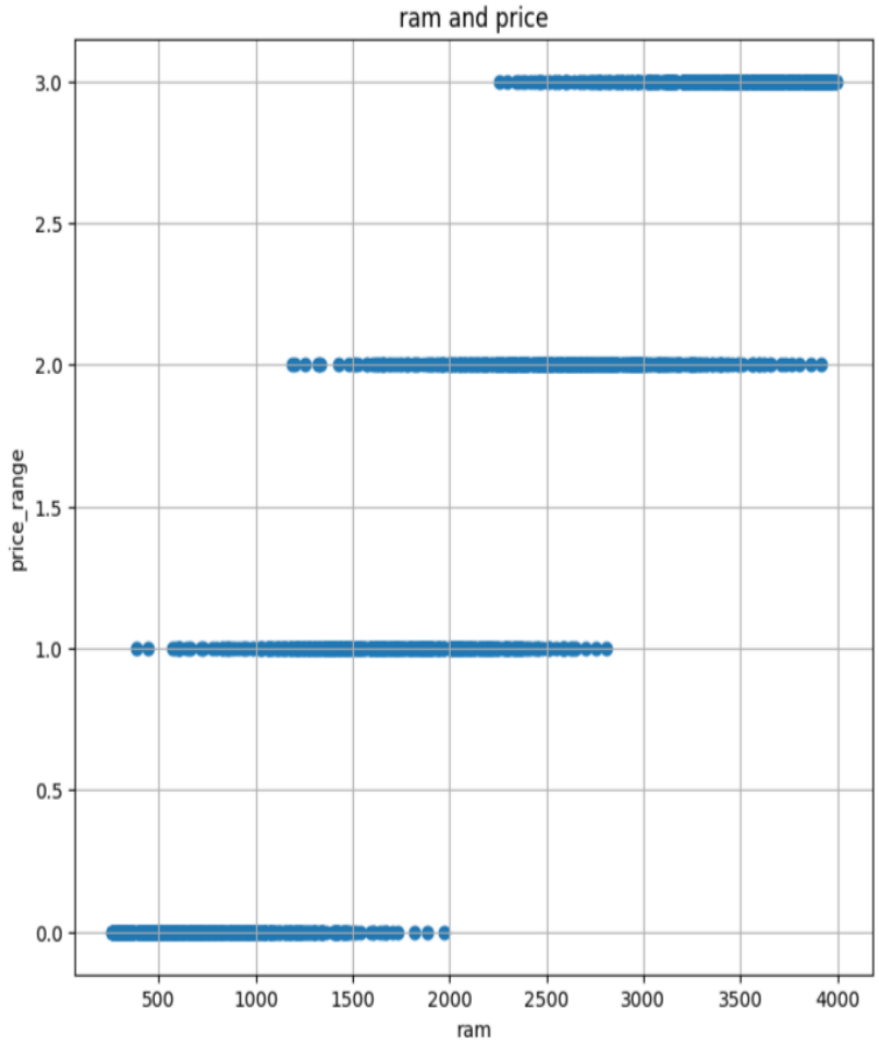
- Fits a Random Forest model to the training data

In [27]:
```
RF.fit(X_train,y_train)
```

Out[27]:
```
RandomForestClassifier
RandomForestClassifier()
```

- Relationship between RAM and price range for mobile phones

In [10]:
```
plt.figure(figsize=(8,8))
plt.scatter(df['ram'],df['price_range'])
plt.title("ram and price")
plt.xlabel('ram')
plt.ylabel('price_range')
plt.grid()
plt.show()
```



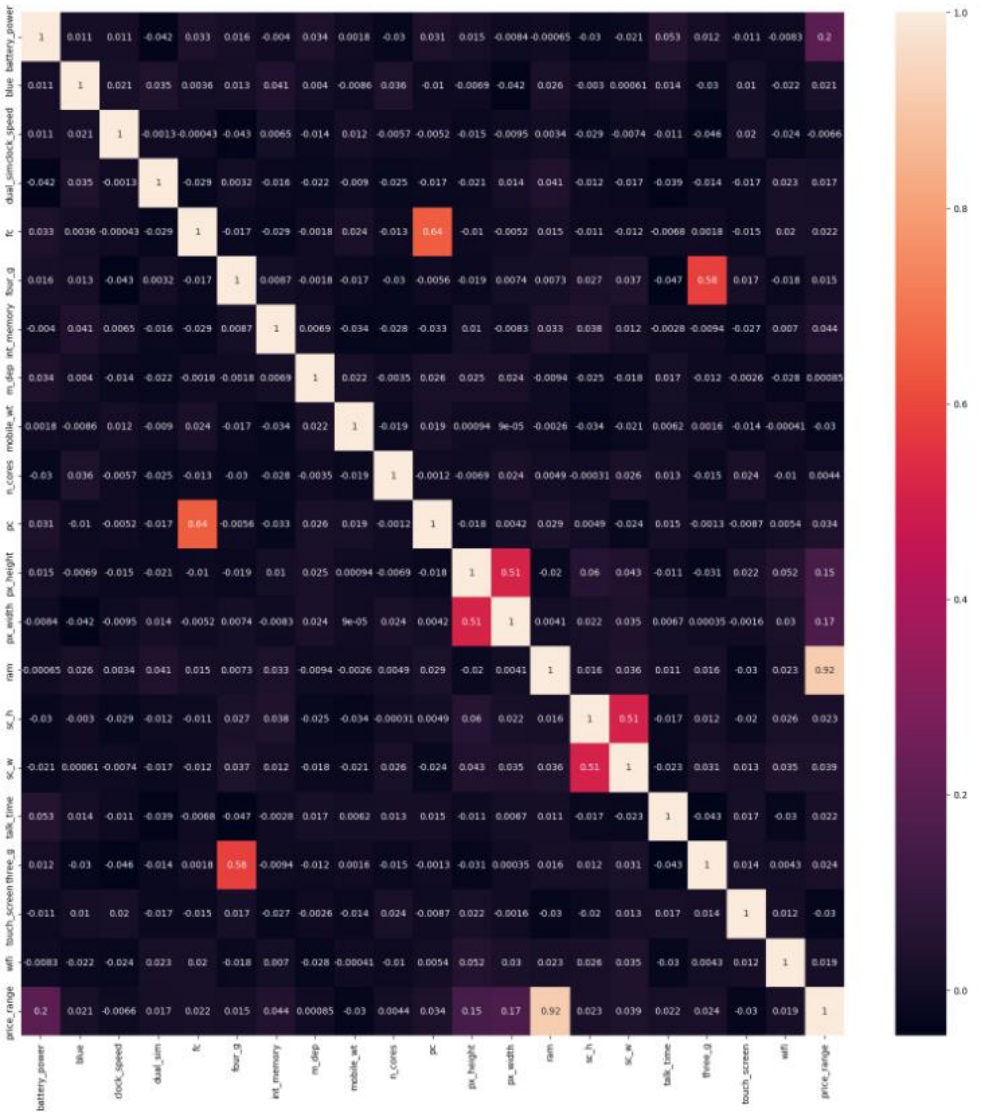
- Heatmap visualization of the correlation matrix

In [11]:
```
import seaborn as sns
```

In [12]:
```
corr = df.corr()
```

In [13]:
```
plt.figure(figsize=(20, 20))
sns.heatmap(corr, annot=True)
```

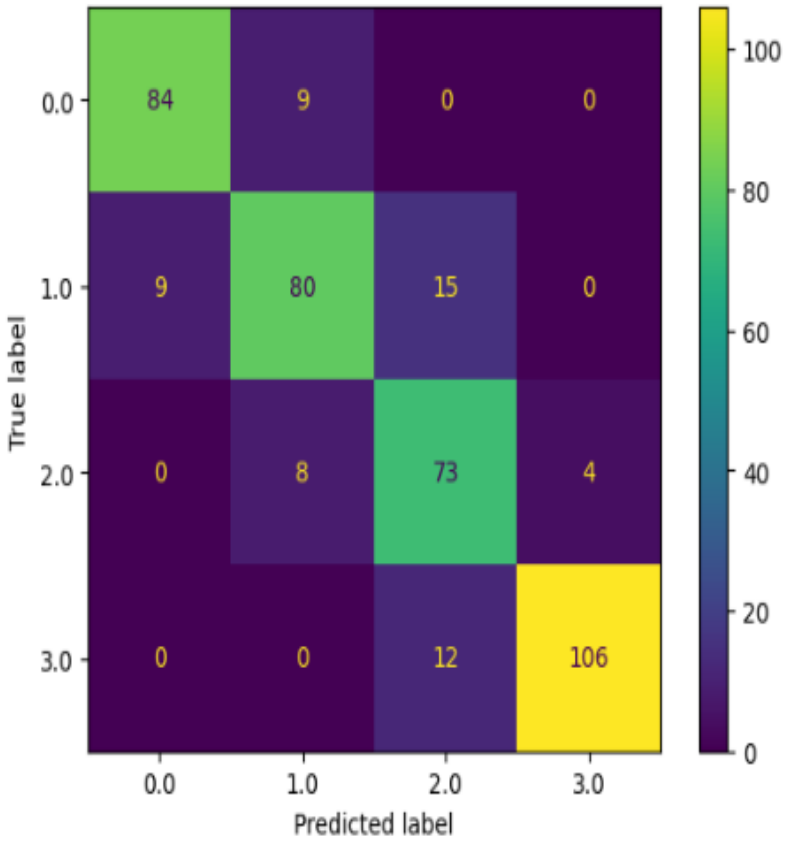
Out[13]:
```
<Axes: >
```



- Confusion matrix from the model's predictions

In [24]:
```
from sklearn.metrics import ConfusionMatrixDisplay
ConfusionMatrixDisplay.from_estimator(model,X_test,y_test)
```

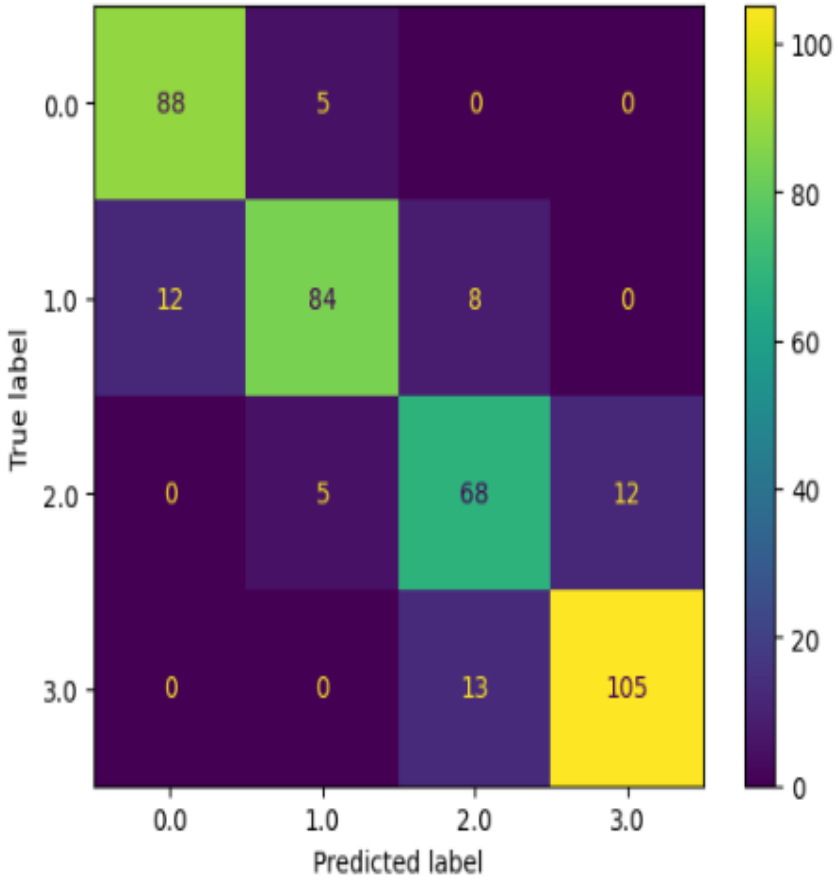
Out[24]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f0165f95d20>
```



- Confusion matrix from the predictions made by the Random Forest classifier

```
In [30]: ConfusionMatrixDisplay.from_estimator(RF,X_test,y_test)

Out[30]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f015f5b32e0>
```



6. CONCLUSION

Through this project, we embarked on a journey to explore and analyse the dynamics of mobile phone price classification, leveraging machine learning techniques. Our endeavour began with meticulous data collection and preparation, where we meticulously curated and cleaned a dataset containing various attributes such as RAM, internal memory, 4G capability, and price range.

Subsequently, we delved into exploratory data analysis (EDA), where we unearthed profound insights into the dataset's intricacies. Employing a plethora of visualization techniques including pie charts, scatter plots, and correlation matrices, we unravelled the underlying patterns and relationships among different features. This analytical phase provided us with invaluable understanding regarding the distribution of features, customer preferences, and potential predictors of price range.

Armed with this newfound knowledge, we ventured into the realm of machine learning, training and evaluating models to predict mobile phone price ranges. Utilizing Support Vector Machine (SVM) and Random Forest classifiers, we harnessed the power of predictive analytics to discern patterns within the dataset and make informed predictions based on features such as RAM, internal memory, and 4G capability.

The performance of these models was meticulously evaluated using accuracy metrics and confusion matrices, affording us a comprehensive understanding of their efficacy in classification tasks. These evaluations underscored the significance of data-driven approaches in elucidating customer behaviour and preferences, illuminating strategic pathways for pricing strategies, marketing campaigns, and product offerings in the fiercely competitive mobile phone market.

In essence, this project encapsulates the fusion of data science and business acumen, illustrating the transformative potential of leveraging data analytics to drive informed decision-making and enhance business outcomes in dynamic industries such as mobile technology.