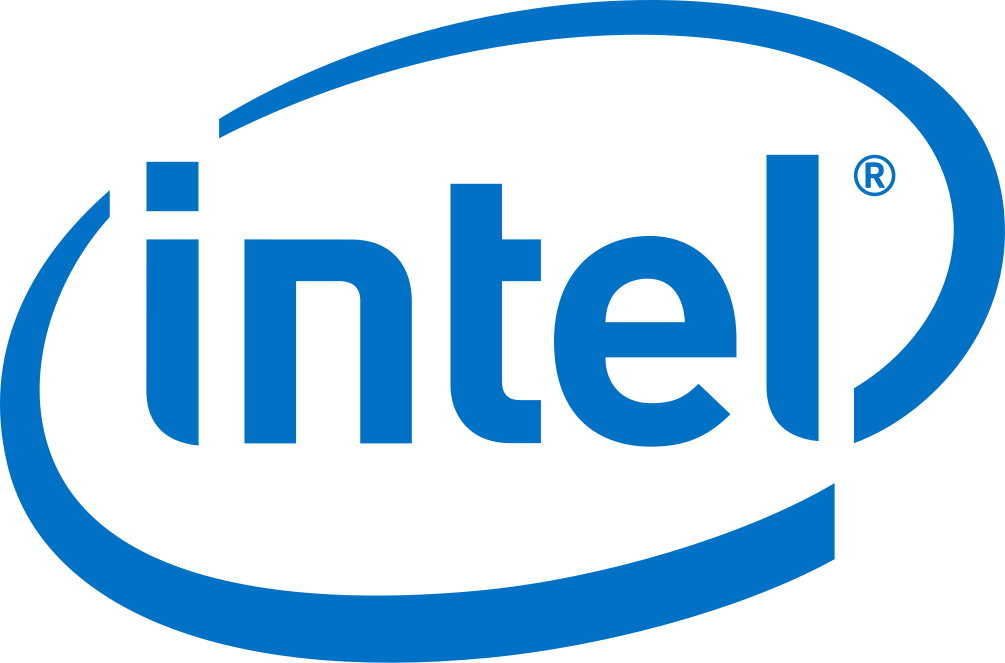
## **Intel Product Review Sentiment Analysis**

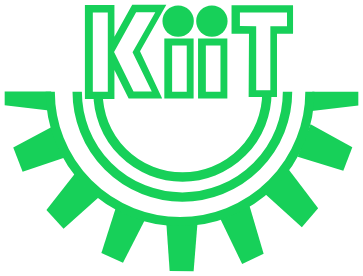
Project under

**Intel Unnati Program 2024 (PS- 11)**

****

By Students of

**Kalinga Institute of Industrial Technology,** Bhubaneswar, Odisha



## **Mentored by: Submitted By:**

## Mr. Debadyut Hazra Rajdeep Thakur (IT-2206113)

## Manish Kumar (IT-2206353)

## Eshita Yadav (IT-2206340)

## Binayak Jha (CSE-22053938)

## **Acknowledgment**

We would like to express our deepest gratitude to our industry mentor, Mr. Debadyut Hazra, for his invaluable guidance, insightful feedback, and continuous support throughout our Intel product sentiment analysis project. His expertise and dedication have been instrumental in the successful completion of this project.

We also extend our sincere thanks to Kalinga Institute of Industrial Technology for providing us with the resources and opportunities to undertake this project. The institution's commitment to fostering academic and professional growth has greatly contributed to our learning experience.

Additionally, we would like to acknowledge the efforts and collaboration of our project team members. Their hard work, dedication, and teamwork have been crucial in achieving our project goals.

Thank you all for your support and contributions.

### **Table of Contents**

**Title Page**

**Acknowledgement**

1. **Introduction**
2. **Data Collection**

2.1 Data Sources

2.2 Tools and Technologies Used

2.3 Methodology

1. **Data Preprocessing**

3.1 Exploratory Data Analysis

1. **Sentiment Analysis Methodology**

4.1 Approach

4.2 Model Selection

4.3 Feature Extraction

1. **Implementation**

5.1 Tools and Libraries

5.2 Model Training

1. **Results and Discussion**
2. **Conclusion**

**References**

**Appendices**

**List of Figures**

| **Figure No.** | **Label** | **Page No.** |
| --- | --- | --- |
| Fig 1.1 | Review Dataset | 6 |
| Fig 2.1 | Class-wise review count before upsampling | 7 |
| Fig 2.2 | Class-wise review count after upsampling | 8 |
| Fig 3.1 | Top 20 most frequent words in Positive Reviews | 9 |
| Fig 3.2 | Word Cloud of Positive Reviews | 9 |
| Fig 3.3 | Top 20 most frequent Bigrams | 10 |
| Fig 6.1 | Model Accuracy Evaluation | 13 |

### 

### 

### **Introduction**

The "Intel Product Sentiment Analysis" project aims to leverage natural language processing (NLP) and machine learning techniques to analyze customer sentiments expressed in reviews of Intel products. By collecting and processing data from various sources, this project seeks to provide actionable insights into customer opinions, preferences, and areas for improvement. The analysis helps Intel better understand market reception, enhance customer satisfaction, and guide strategic decision-making for future product development.

### 

### **Data Collection**

**Data Sources**: Amazon

**Tools and Technologies Used:** Python, Selenium, ChromeDriver

**Methodology:**The data collection process involves setting up a Python environment with Selenium and ChromeDriver for web scraping. Given that the links to Amazon review pages for various Intel processors (10th, 11th, and 12th generations of Core i3, i5, i7, and i9) are dynamic, Selenium is used to navigate these URLs and extract review data. The data extraction process includes collecting review text, ratings, dates, and reviewer information. Scrolling and pagination techniques are implemented to ensure comprehensive data extraction from multiple pages. The reviews are then categorized as positive, negative, or neutral based on their content.

A total of 5,496 reviews have been scraped and saved in a structured CSV format for further analysis.

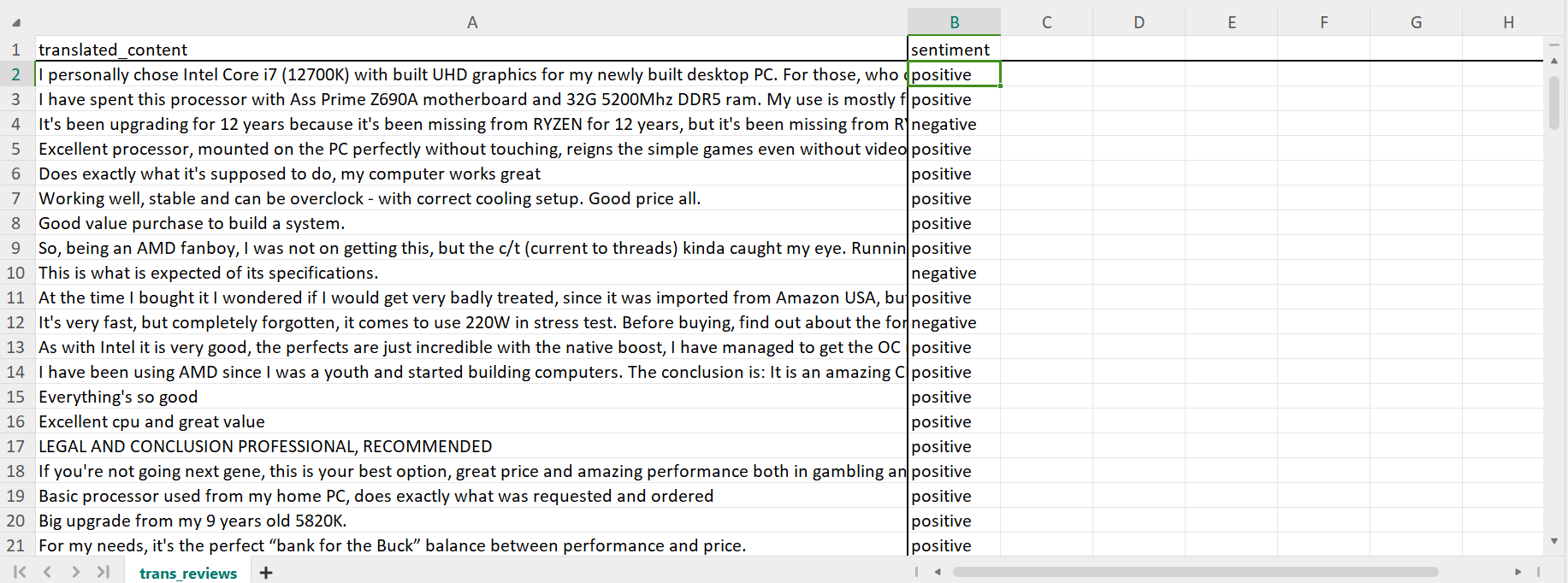


Fig 1.1- Review Dataset

**Data Preprocessing**

The data preprocessing step involves ensuring data integrity and addressing class imbalance.

Upon inspecting the dataframe,no missing values were found. However, the review distribution revealed a significant class imbalance with 2567 positive reviews, 1673 neutral reviews, and 1256 negative reviews.

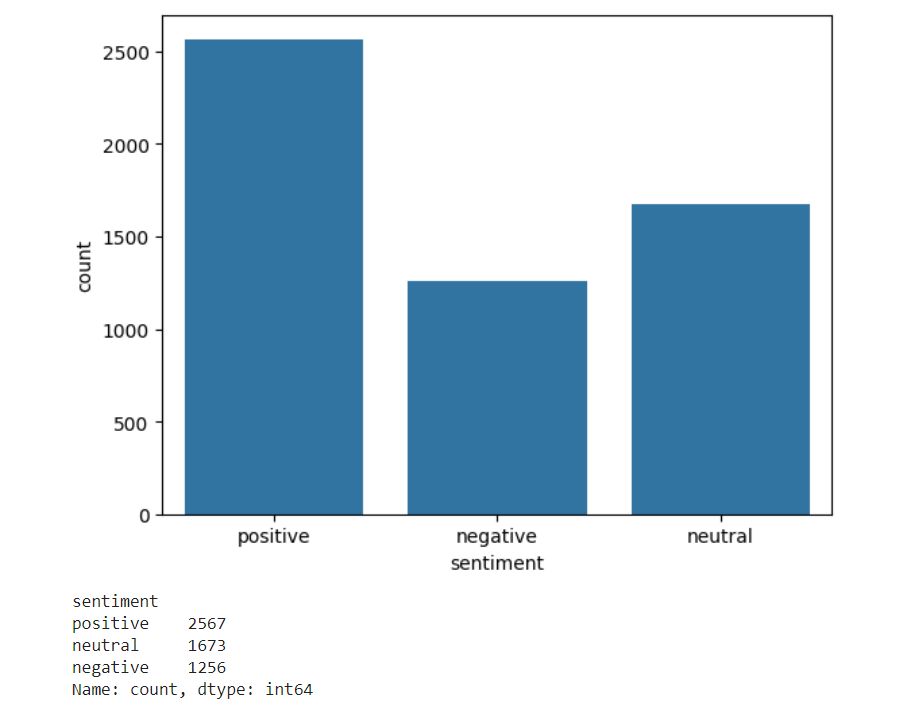


Fig 2.1- Class-wise review count before upsampling

To mitigate this imbalance, upsampling was applied to the minority classes (neutral and negative), bringing each sentiment category to an equal count of 2567 reviews.

The upsampled text data in the dataframe was stemmed to reduce words to their root forms, optimizing it for further analysis.

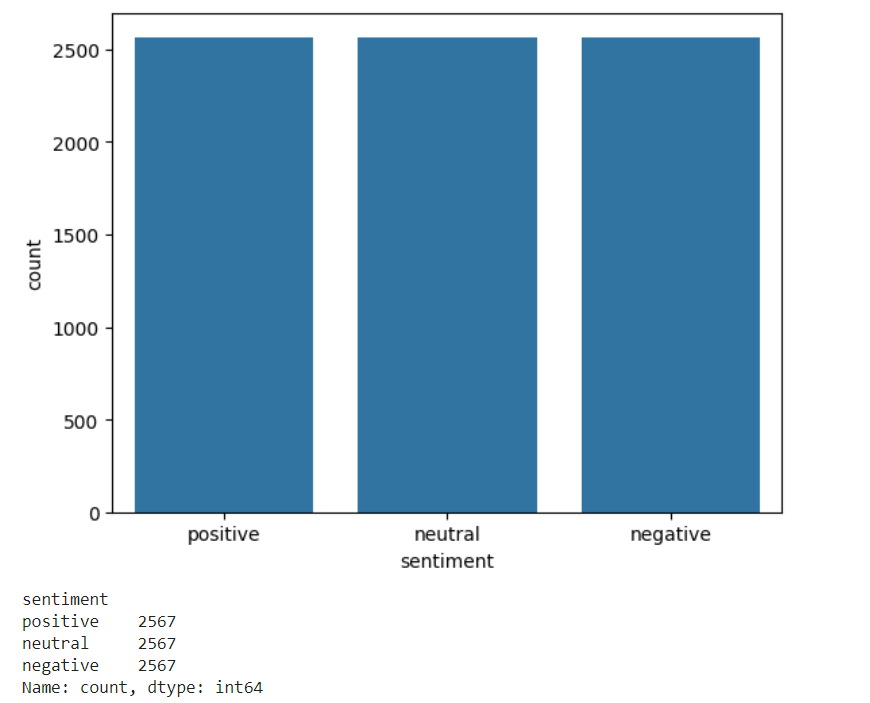
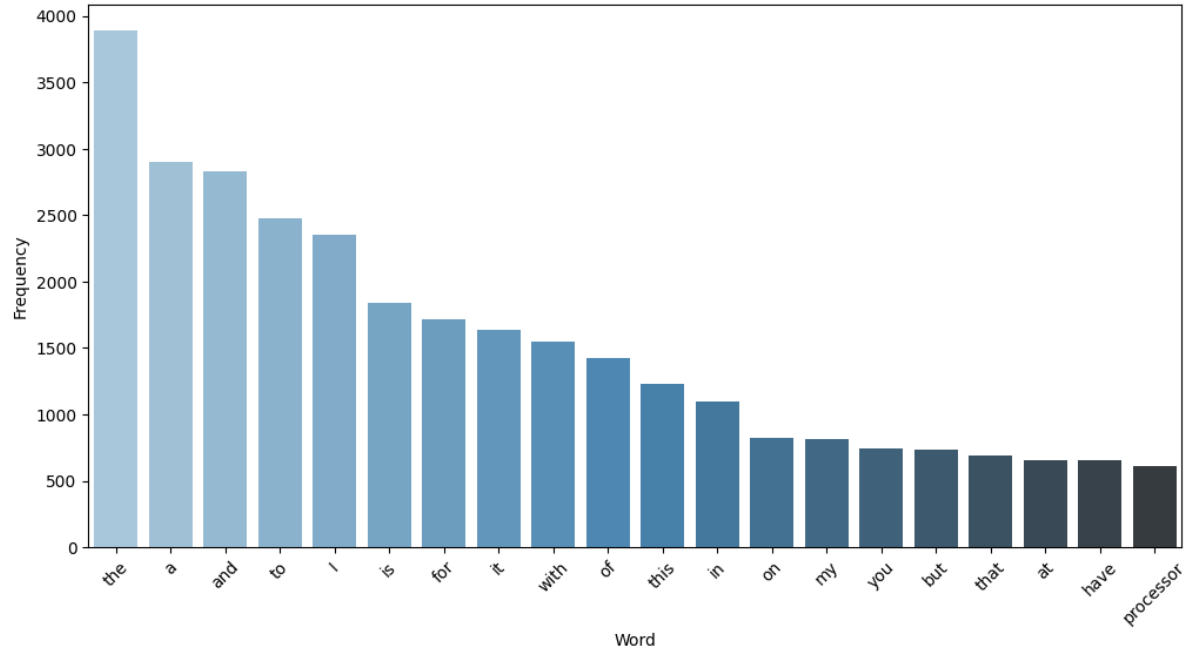


Fig 2.2- Class-wise review count after upsampling

**Exploratory Data analysis (EDA):** Different types of distributions are plotted, such as the number of words in positive, neutral, and negative reviews, as well as the top 20 most frequent words.

Fig 3.1- Top 20 most frequent words in Positive Reviews

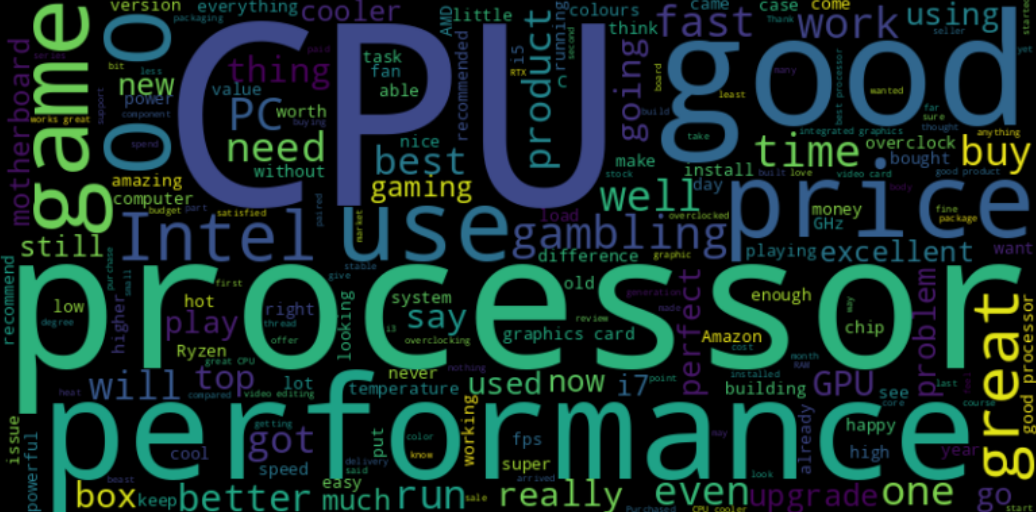


Fig 3.2- Word Cloud of Positive Reviews

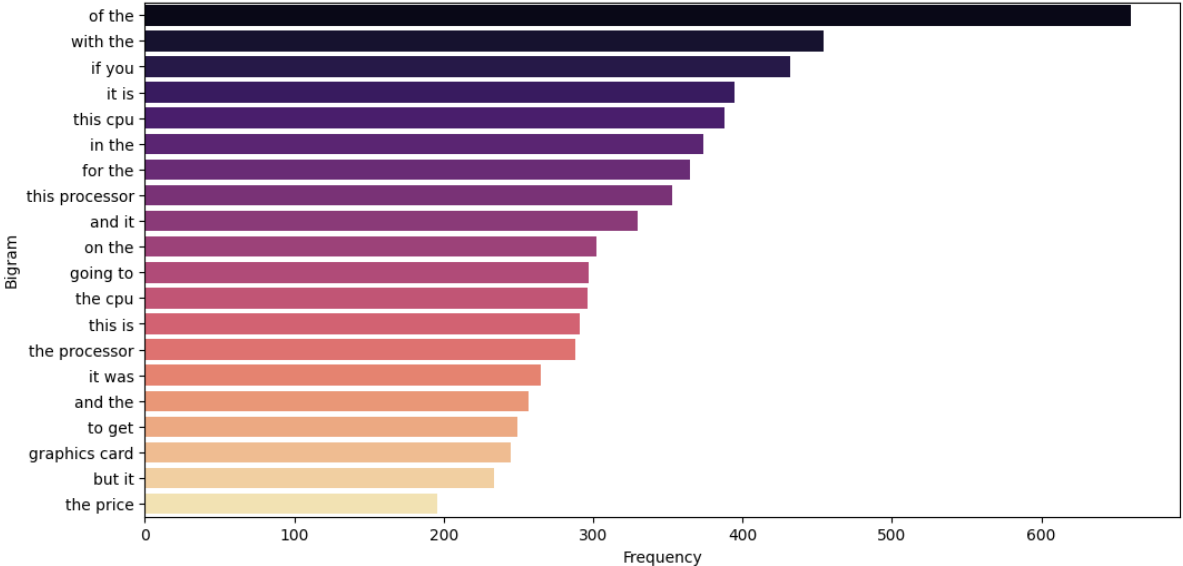


Fig 3.3- Top 20 most frequent Bigrams

### **Sentiment Analysis Methodology**

### **Approach**: In this project logistic regression was utilized for its effectiveness in binary classification tasks, where the goal is to categorize reviews into positive, neutral, or negative sentiments based on extracted features from text data.

### **Model Selection**: Logistic regression was selected as the model due to several reasons. It is well-suited for binary classification tasks, provides interpretable results in terms of probabilities for each class, and performs efficiently with moderate to large datasets. Logistic regression is also less prone to overfitting compared to more complex models, making it suitable for tasks where the decision boundary between classes is relatively straightforward, such as sentiment analysis of textual data.

### **Feature Extraction**: Feature extraction methods employed include TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF is utilized to transform text data into numerical features that represent the importance of words in reviews relative to the entire dataset. This method helps in capturing the distinguishing characteristics of words in different sentiment categories, thereby improving the model's ability to classify reviews accurately based on their textual content.

**Implementation**

**Tools and Libraries:** The model was implemented using various software tools and libraries, including Python as the programming language, along with essential libraries such as NLTK (Natural Language Toolkit), scikit-learn for machine learning tasks. NLTK was specifically used for text preprocessing tasks like stemming and stopword removal, while scikit-learn provided tools for data preprocessing, model training, and evaluation.

**Model Training**: The training process involved several key steps. First, the dataset was split into training and testing sets using a specified ratio (e.g., 80% training, 20% testing) to ensure the model's performance could be evaluated accurately on unseen data. Then the textual data has been converted into numerical data using TFIDF Vectorizer. The Logistic Regression model has then been run on the dataset. Model accuracy has been evaluated first on the trained data then on the test data.

**Result and Discussion**

The proposed model predicts the data correctly with an accuracy score of 98.49% for training data and 97.46% for test data.

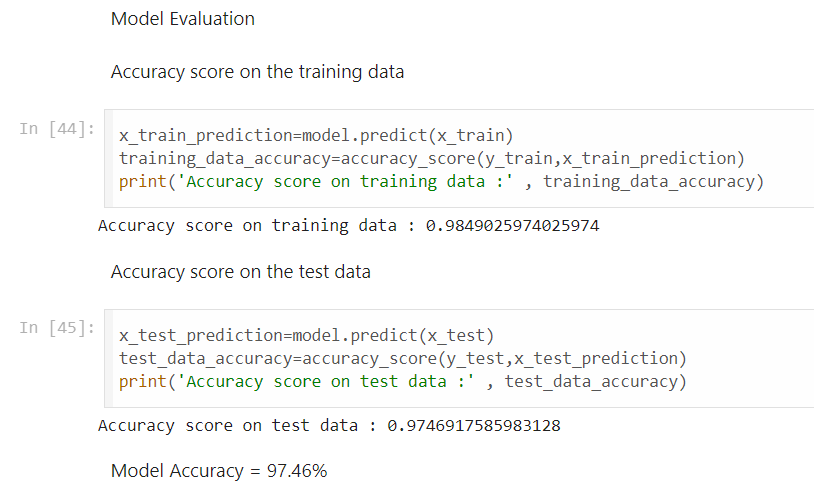


Fig 6.1- Model AccuracyEvaluation

**Conclusion**

The Intel Product Sentiment Analysis project successfully leveraged natural language processing and machine learning techniques to evaluate customer sentiments towards Intel products. By analyzing customer reviews, we were able to categorize sentiments into positive, neutral, and negative classes, providing valuable insights into consumer perceptions. The project highlights the effectiveness of automated sentiment analysis in understanding customer feedback, which can aid Intel in enhancing product quality and customer satisfaction.

**References**

GeeksForGeeks- [**https://www.geeksforgeeks.org/**](https://www.geeksforgeeks.org/)

Java Point- [**https://www.javatpoint.com/**](https://www.geeksforgeeks.org/)

**Appendices**

**Product Links:**

<https://www.amazon.in/Intel-i7-12700K-Desktop-Processor-Unlocked/product-reviews/B09FXNVDBJ/> <https://www.amazon.in/Intel-Generation-Desktop-Processor-Warranty/product-reviews/B09MDFH5HY/> <https://www.amazon.in/Intel-i7-13700K-Desktop-Processor-P-cores/product-reviews/B0BCF57FL5/> <https://www.amazon.in/Intel-i7-11700K-Processor-Integrated-Graphics/product-reviews/B08X6ND3WP/> <https://www.amazon.in/Intel%C2%AE-i7-14700KF-Desktop-Processor-P-cores/product-reviews/B0CGJC178L/> <https://www.amazon.in/Intel%C2%AE-i7-9700K-Processor-Cache-Socket/product-reviews/B07HHN6KBZ/> <https://www.amazon.in/Intel-i7-10700K-Desktop-Processor-Unlocked/product-reviews/B086ML4XSB/> <https://www.amazon.in/Intel%C2%AE-CoreTM-i7-10700-Processor-Cache/product-reviews/B086MG1C7C/><https://www.amazon.in/Intel%C2%AE-CoreTM-i3-10100-Processor-Cache/product-reviews/B086MMRW87/><https://www.amazon.in/Intel%C2%AE-CoreTM-i3-12100-Processor-Cache/product-reviews/B09MDDX29R/><https://www.amazon.in/Intel-i3-10100F-Generation-LGA1200-Processor/product-reviews/B08LKJPR5X/> <https://www.amazon.in/Intel-BX8071512100F-INTEL-I3-12100F-DESKTOP/product-reviews/B09NPJX7PV/> <https://www.amazon.in/Intel%C2%AE-CoreTM-i5-10400F-Processor-Cache/product-reviews/B086MHSTWN/> <https://www.amazon.in/Intel-i5-11600KF-Desktop-Processor-Unlocked/product-reviews/B08X62BTJD/> <https://www.amazon.in/Intel%C2%AE-i5-9400F-Processor-Cache-Socket/product-reviews/B07MRCGQQ4/> <https://www.amazon.in/Intel-Generation-Desktop-Processor-Warranty/product-reviews/B09MDGKQLY/> <https://www.amazon.in/Intel-Unlocked-LGA1151-Desktop-Processor/product-reviews/B07HHLX1R8/>