



**Faculty of Engineering & Technology**  
Department of Electrical & Computer Engineering

## **Accelerated DLRM-based E-commerce Recommendation System**

### **Prepared By:**

Ibraheem Alyan    1201180  
Mohammad Abu-Shelbaia    1200198  
Nidal Zabade    1200153

### **Supervised By:**

Dr. Ahmed Shawahna

A Graduation Project submitted to the Department of Electrical and  
Computer Engineering in partial fulfillment of the requirements for the  
degree of B.Sc. in Computer Engineering

Birzeit

December 19, 2023

# Abstract

# المستخلص

أكتب هنا

# Table of Contents

English Abstract	I
Arabic Abstract	II
Table of Contents	III
List of Tables	V
List of Figures	VI
<b>1 Introduction and Motivation</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Problem Statement . . . . .	2
1.3 Report Outline . . . . .	2
<b>2 Background</b>	<b>3</b>
2.1 Transformer . . . . .	3
2.1.1 Model Architecture . . . . .	3
2.1.2 Scaled Dot-Product Attention . . . . .	3
2.1.3 Multi-Head Attention . . . . .	3
2.1.4 Self-Attention and Multi-Head Self-Attention . . . . .	3
2.1.5 Feed Forward Network . . . . .	3
2.2 Vision Transformer (ViT) . . . . .	3
2.3 Lightweight ViT . . . . .	3
<b>3 Literature Review—ViT Acceleration Techniques</b>	<b>4</b>
3.1 Pruning . . . . .	4
3.2 Quantization . . . . .	4
3.3 Low-Rank Approximation . . . . .	4
3.4 Knowledge Distillation . . . . .	4

3.5	Lightweight ViT . . . . .	4
3.6	Transformer Acceleration on Hardware . . . . .	4
<b>4</b>	<b>Proposed Work</b>	<b>5</b>
<b>5</b>	<b>Project Plan</b>	<b>6</b>
<b>6</b>	<b>Conclusion and Future Work</b>	<b>7</b>
	<b>Bibliography</b>	<b>8</b>

# List of Tables

# List of Figures

# Chapter 1

## Introduction and Motivation

### Contents

1.1	Motivation . . . . .	1
1.2	Problem Statement . . . . .	2
1.3	Report Outline . . . . .	2

### 1.1 Motivation

The exponential growth of e-commerce has introduced an enormous amount of choice, where consumers face overwhelming product options. To address this challenge, personalized recommendation systems have become essential for enhancing the shopping experience, and increasing the conversion rate for any e-commerce platform.

In contrast to conventional collaborative filtering[1], content-based[1], or popularity-based recommendation systems, our AI-based solution offers distinct advantages. Firstly, AI makes it possible to provide per-user personalized recommendations, which are tailored to their unique preferences and behaviors, enhancing user engagement and satisfaction. AI systems can also intelligently recommend comparable or complementary products or content to increase revenue through cross-selling. Furthermore, AI takes into account the impressions and interactions of users with items, allowing for a more dynamic and accurate understanding of user preferences. Using AI leads to improved recommendation accuracy and relevancy, leading to increased conversion rates and business growth.

Statistics from different use cases of recommendation systems:

- On average, an intelligent recommender system delivers a 22.66% lift in conversions rates [2] for web products.
- IKEA experienced a 30% increase in click-through rate, 2% surge in average order value [3] using Google Recommendations AI [4]
- Lotte Mart increased new product purchases by 1.7x [5] using Amazon Personalize [6]

In summary, our project's motivation is elevating the e-commerce experience, driving



business success, and harnessing cutting-edge AI technologies to create a recommendation system that is both high-performing and scalable.

## 1.2 Problem Statement

the process of building the solution is mainly two parts:

- First, designing a personalized recommendation system that covers what traditional collaborative filtering, content-based, or popularity-based systems cannot achieve.
- Second, deploying and automating the solution, including, data cleaning, data storage, and model deployment processes, and ensuring a production-ready and scalable system.

## 1.3 Report Outline

TODO

# Chapter 2

## Background

### Contents

---

<b>2.1</b>	<b>Transformer . . . . .</b>	<b>3</b>
2.1.1	Model Architecture . . . . .	3
2.1.2	Scaled Dot-Product Attention . . . . .	3
2.1.3	Multi-Head Attention . . . . .	3
2.1.4	Self-Attention and Multi-Head Self-Attention . . . . .	3
2.1.5	Feed Forward Network . . . . .	3
<b>2.2</b>	<b>Vision Transformer (ViT) . . . . .</b>	<b>3</b>
<b>2.3</b>	<b>Lightweight ViT . . . . .</b>	<b>3</b>

---

## 2.1 Transformer

### 2.1.1 Model Architecture

### 2.1.2 Scaled Dot-Product Attention

### 2.1.3 Multi-Head Attention

### 2.1.4 Self-Attention and Multi-Head Self-Attention

### 2.1.5 Feed Forward Network

## 2.2 Vision Transformer (ViT)

## 2.3 Lightweight ViT

# Chapter 3

## Literature Review—ViT Acceleration Techniques

### Contents

---

3.1	Pruning . . . . .	4
3.2	Quantization . . . . .	4
3.3	Low-Rank Approximation . . . . .	4
3.4	Knowledge Distillation . . . . .	4
3.5	Lightweight ViT . . . . .	4
3.6	Transformer Acceleration on Hardware . . . . .	4

---

### 3.1 Pruning

### 3.2 Quantization

### 3.3 Low-Rank Approximation

### 3.4 Knowledge Distillation

### 3.5 Lightweight ViT

### 3.6 Transformer Acceleration on Hardware

## Chapter 4

### Proposed Work

# Chapter 5

## Project Plan

## Chapter 6

### Conclusion and Future Work

# Bibliography

- [1] Nvidia Glossary, “Recommendation System.” <https://www.nvidia.com/en-us/glossary/data-science/recommendation-system/>. Accessed: 2023-12-19.
- [2] Salesforce Marketing Cloud, “Predictive intelligence benchmark report.” <https://brandcdn.exacttarget.com/sites/exacttarget/files/deliverables/etmc-predictiveintelligencebenchmarkreport.pdf>, 2014.
- [3] Google Cloud Summit, “Ikea’s approach to building a powerful recommendations engine.” by Google Cloud Events <https://www.youtube.com/watch?v=PyjC0wRRtBg>, 2021.
- [4] Google Cloud, “Google Recommendations AI.” <https://cloud.google.com/recommendations>. Accessed: 2023-11-9.
- [5] Sungoh Park and Kyoungtae Hwang, “Increasing customer engagement and loyalty with personalized coupon recommendations using Amazon Personalize,” *AWS Machine Learning Blog*, 2020.
- [6] AWS, “AWS Personalize.” <https://aws.amazon.com/personalize/>, accessed: 2023-10-4.