T.C.
YILDIZ TECHNICAL UNIVERSITY
FACULTY OF MECHANICAL ENGINEERING
DEPARTMENT OF INDUSTRIAL ENGINEERING

UNDERGRADUATE GRADUATION THESIS WRITING GUIDE

Muhammed DİLBER

Advisor
Prof. Dr. Alev TAŞKIN

UNDERGRADUATE GRADUATION THESIS

JUNE, 2025

TABLE OF CONTENTS

LIST OF SYMBOL

A, B	: Semantic vectors representing two different courses or a query and a course
A·B	: Dot product between vectors A and B
$\|\|A\|\|, \|\|B\|\|$	: Euclidean norms (magnitudes) of vectors A and B
θ	: Angle between vectors A and B
$cos(\theta)$	: Cosine of the angle between A and B
Si	: Similarity score between user query vector and course vector vi
u	: User query semantic vector
vi	: Semantic vector of course i
n	: Number of dimensions after TruncatedSVD

LIST OF ABBREVIATION

BERT	: Bidirectional Encoder Representations from Transformers
CRF	: Conditional Random Fields
Django	: Python-based Web Framework
FSLSM	: Felder-Silverman Learning Style Model
GNN	: Graph Neural Network
ID	: Identifier
LSA	: Latent Semantic Analysis
NLP	: Natural Language Processing
PostqreSQL	: Relational Database Management System
SBERT	: Sentence-BERT (Sentence Embeddings using BERT)
SDLC	: Software Development Life Cycle
YTU	: Yıldız Technical University

LIST OF FIGURE

LIST OF TABLE

ABSTRACT

Muhammed DİLBER

Industrial Engineering Department Graduation Thesis

Advisor

Prof. Dr. Alev TAŞKIN

In recent years, it has been observed that university students experience various difficulties in course selection processes and that this process leads to erroneous or unsatisfactory results for many students. Common problems include the fact that course contents do not appeal to students' interests, that students have difficulty in identifying the courses they want to specialize in, and that they do not want to devote enough time to the decision process.

The aim of this study is to develop a recommendation system that enables students at Yıldız Technical University Department of Industrial Engineering to find elective courses that match their interests in a faster and more logical way. The developed system analyses the course contents with natural language processing (NLP) methods and recommends the courses that best match the interests expressed by the students in writing.

The project was developed for Yıldız Technical University (YTU) Industrial Engineering students to enable them to choose the right elective course in the field they want to improve themselves. These textual data are then embedded into high-dimensional semantic vectors using the SBERT (Sentence-BERT) model. To reduce computational complexity and remove redundant features, Latent Semantic Analysis (LSA) is applied for dimensionality reduction. The user input is used to sort the courses of the relevant elective category using dot product.

The interface is developed in Django using Python. The system allows students to explore elective courses, receive personalized recommendations and review detailed course information. The recommendation system uses text-based course descriptions, semantic matching with SBERT and LSA models, and ranks appropriate courses based on user input.

Keywords: Course Recommendation, SBERT, LSA, Semantic Similarity, Natural Language Processing, Django.

ÖZET

Muhammed DİLBER

Endüstri Mühendisliği Bölümü Bitirme Çalışması

Danışman Öğretim Üyesi
Prof. Dr. Alev TAŞKIN

Son yıllarda üniversite öğrencilerinin ders seçim süreçlerinde çeşitli zorluklar yaşadıkları ve bu sürecin birçok öğrenci için hatalı veya tatmin edici olmayan sonuçlara yol açtığı gözlemlenmektedir. Ders içeriklerinin öğrencilerin ilgi alanlarına hitap etmemesi, öğrencilerin uzmanlaşmak istedikleri dersleri belirlemekte zorlanmaları ve karar sürecine yeterince zaman ayırmak istememeleri yaygın sorunlar arasında yer almaktadır.

Bu çalışmanın amacı, Yıldız Teknik Üniversitesi Endüstri Mühendisliği Bölümü öğrencilerinin ilgi alanlarına uygun seçmeli dersleri daha hızlı ve mantıklı bir şekilde bulmalarını sağlayan bir öneri sistemi geliştirmektir. Geliştirilen sistem, ders içeriklerini doğal dil işleme (NLP) yöntemleri ile analiz etmekte ve öğrencilerin yazılı olarak ifade ettikleri ilgi alanlarına en uygun dersleri önermektedir.

Proje, YTÜ Endüstri Mühendisliği öğrencilerinin kendilerini geliştirmek istedikleri alanda doğru seçmeli dersi seçebilmeleri için geliştirilmiştir. Bu metinsel veriler daha sonra SBERT modeli kullanılarak yüksek boyutlu anlamsal vektörlere gömülmüştür. Hesaplama karmaşıklığını azaltmak ve gereksiz özellikleri ortadan kaldırmak için, boyut azaltma amacıyla Gizli Anlamsal Analiz (LSA) uygulanır. Kullanıcı girdisi, ilgili seçmeli kategorideki dersleri nokta çarpımı kullanarak sıralamak için kullanılır.

Arayüz Python kullanılarak Django'da geliştirilmiştir. Sistem, öğrencilerin seçmeli dersleri keşfetmelerine, kişiselleştirilmiş öneriler almalarına ve ayrıntılı ders bilgilerini incelemelerine olanak tanır. Öneri sistemi, metin tabanlı ders açıklamalarını, SBERT ve LSA modelleri ile anlamsal eşleştirmeyi kullanır ve kullanıcı girdisine göre uygun dersleri sıralar.

Anahtar Kelimeler: Ders Öneri Sistemi, SBERT, LSA, Anlamsal Benzerlik, Doğal Dil İşleme, Django.

INTRODUCTION

1.1. General Introduction to the Topic

In recent years, it has been observed that university students experience various difficulties in course selection processes and that this process leads to erroneous or unsatisfactory results for many students. Common problems include the fact that course contents do not appeal to students' interests, that students have difficulty in identifying the courses they want to specialize in, and that they do not want to devote enough time to the decision process. As a student who has personally experienced these problems, this is where I got the idea to develop this study.

The rapid development of NLP and recommender systems in recent years provides important opportunities to understand user needs and provide individualized recommendations. In this context, we aim to develop a recommendation system that allows students in YTU Industrial Engineering undergraduate program to analyse course content, see courses in similar categories, and find the courses they want as soon as possible according to their interests. The recommendation system uses text-based course descriptions, semantic matching with SBERT and LSA models, and ranks appropriate courses based on user input.

1.2. Problem Definition

For undergraduate students, the courses in the curriculum, the content and scope of the courses are very important. What is more important is that they can take the courses they want. However, the excessive number of elective courses, the high number of courses, knowing the course names but not having an idea about the content are among the difficulties faced by the students.

For YTU Industrial Engineering students, analyzing and comparing the courses one by one is quite laborious and time consuming. This situation causes students to choose courses that they are not interested in and causes loss of motivation. To prevent such situations, there is a need to develop a system that can make personalized and meaning-based course recommendations according to the needs of students.

1.3. Objective of the Study

The aim of this study is to develop a recommendation system that enables students at Yıldız Technical University Department of Industrial Engineering to find elective courses that match their interests in a faster and more logical way. The developed system analyses the course contents with NLP methods and recommends the courses that best match the interests expressed by the students in writing.

1.4. Methodological Approach

In this study, the elective courses offered within the Department of Industrial Engineering at Yıldız Technical University are categorized into four main groups: Vocational Elective A, Vocational Elective B, Social Elective and University Vocational Elective. Comprehensive course information for each group was retrieved from the university's Bologna Information System. The data included course name, course code, prerequisites, semester, course language, course level, course category, course unit, course objectives, course content, recommended textbooks, learning outcomes, weekly topics and assessment methods. This data was stored in a database.

The elective courses in each group were then manually reviewed and categorized according to their similarities. These category labels were then assigned to each lesson to facilitate more meaningful recommendation matching later in the process.

For the NLP phase, the following text fields were selected to represent each course: course name, course objective, course content, learning outcomes and weekly topics. The information in these fields constitutes the content of the course.

The SBERT model was used to transform these combined course texts into high-dimensional semantic embeddings. LSA was applied as a dimensionality reduction technique to improve performance and reduce noise. After this preprocessing, the similarity between the student input (expressed as free text) and the lesson vectors was calculated using cosine similarity. The courses with the highest similarity scores were presented as personalized recommendations.

1.5. Scope and Limitations of the Study

The scope of the study is limited to four different types of elective courses (University Vocational Elective, Social Elective, Vocational A, Vocational B) in YTU Industrial Engineering Department. Course contents were retrieved from the university's bologna website and students' interests were collected through text input. The model is based on content-based filtering only; user behavior, past preferences, or academic performance data are not included in the analysis.

This recommendation system was developed for YTU Industrial Engineering students and is not intended to be generalized to other departments. However, since social elective courses are common to all departments, only this part can be used by other department students.

1.6 Place in the Literature and Contribution

In this sub-topic, some papers on the use of recommender systems in education are discussed. The contents of these articles are analyzed. These papers explore various approaches such as collaborative filtering, content-based models, hybrid architectures and chatbots, and neural networks. The selected papers include applications such as personal course recommendations and adaptive educational platforms from learning resource recommendations.

These papers have been selected to address how recommender systems in education have been applied in different settings and how to improve student engagement, performance and personal learning experiences.

Despite the growing popularity of recommender systems in education, some needs remain. Most existing systems rely heavily on static student profiles or simple filtering techniques. Few studies rely on rich textual content such as direct course descriptions, learning outcomes or weekly topics.

In addition, there has been limited interest in elective course recommendation systems specifically designed for higher education that integrate students' interests with natural language understanding techniques.

Therefore, this paper aims to address this gap by developing a course recommendation system targeting elective courses in Industrial Engineering at YTU, using SBERT for semantic vectorization and LSA for dimensionality reduction. The proposed system differentiates itself by providing a deeper understanding of students' course interests and providing personalized recommendations.

1.7 Structure of the Report

This thesis is organized in five main chapters with the aim of developing a personalized elective course recommendation system for the students of YTU, Department of Industrial Engineering.

Chapter 1 introduces the motivation of the work, defines the problem and gives an overview of the methodology.

Chapter 2 presents a literature review on the use of recommender systems, sentence embedding, recommender models and dimensionality reduction techniques in NLP.

Chapter 3 describes the architectural design of the developed recommender system, including the data structure, embedding, LSA implementation and user interface components.

Chapter 4 describes the implementation details, data collection and preparation, categorization of student inputs, LSA, cosine product, dot product methods. In addition, multiple case studies are presented to illustrate the system performance.

Chapter 5 summarizes the thesis, discusses the results, highlights contributions and limitations, and provides suggestions for future work.

LITERATURE REVIEW

2.1. Importance of Recommender Systems in Education

Recommender systems are generally structures for recommending products or services to users using some techniques. In 1990 and 1994, two different studies came to the agenda for the first time with the naming of digital libraries (Goldberg, Nichols, Oki, & Terry, 1992; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994). It is used in many systems and infrastructures such as radio programs, e-commerce sites, dating applications, social media channels. In structures where the content is very large, it responds to the need to filter the content according to the user's personal interests. These are systems that provide solutions to the need to deliver smaller data sets to the user by reducing data structures with large content to subsets based on the user's interests. (Goldberg, Nichols, Oki, & Terry, 1992; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994).

Basically, there are four important types of recommendation systems:

- Collaborative Filtering Systems
- Content-Based Filtering Systems
- Popularity-Based Recommendation Systems
- Hybrid Recommendation Systems (Roy & Dutta, 2022)

2.2. Semantic Similarity and Sentence Embedding (SBERT, BERT, and Other Methods)

<u>Text Similarity and Its Evolution</u>

Text similarity refers to the semantic similarity of two texts or expressions. Especially today, it plays an important role in personalizing user preferences in recommendation systems. Early similarity techniques mainly relied on comparing word frequencies or exact matches between documents. However, these shallow methods (e.g., bag-of-words or TF-IDF) fail to capture the semantic closeness between two texts. For example, two descriptions such as “basics of machine learning” and “introduction to artificial intelligence” may be conceptually related, but previous similarity techniques may treat them as completely different. This is where the need for more advanced similarity and matching strategies arises. (De Boom et al., 2015)

Semantic Similarity and Embeddings

Semantic similarity measures how similar two words or texts are in meaning or scope. Semantic similarity is based on embeddings. These embeddings allow models to compare not only word overlaps but also underlying concepts. For example, “neural networks” and “deep learning models” may have different vocabularies but closely related semantic content. (De Boom et al., 2015)

Word Embeddings and Sentence Embeddings

Examples of traditional word embedding methods are Word2Vec and GloVe. These methods represent words as fixed vectors. The problem with this logic is that it cannot capture the meaning of words according to where they are used in sentences. For example, homophones have different meanings depending on their usage in a sentence. Without a model that looks at the whole sentence, it is not possible to distinguish this. Because the same vector is assigned to the word regardless of its meaning. (Mikolov et al., 2013). BERT (Bidirectional Encoder Representations from Transformers), developed to overcome this limitation, handles polysemy more effectively by representing words according to their context. (Devlin et al., 2019) However, since BERT is not efficient for direct sentence similarity computations, the SBERT) model was developed. SBERT processes the output of BERT with siamese and triplet network structures, transforming sentences into fixed-length vectors and enabling similarity computations (e.g., cosine similarity) between these vectors. (Reimers & Gurevych, 2019)

Cosine Similarity and Dot Product

Cosine similarity is a metric used to measure the similarity between two vectors in an inner product space. It measures how closely two vectors, regardless of their magnitude, are aligned by calculating the cosine of the angle between them. The formula for cosine similarity is as follows: Cosine Similarity (A, B) = $(A \cdot B) / (||A|| * ||B||)$ (Luo et al., 2017).

The dot product measures both the directional similarity and the magnitude of the vectors. It is especially used when the magnitude of the vectors is also important. The formula for the dot product is as follows: Dot product(a,b) = $a \cdot b = |a||b|\cos\theta$. (Luo et al., 2017).

2.3. Dimensionality Reduction Using LSA(Luo et al., 2017).

One of the biggest challenges of data science is multidimensional features. Multidimensional features incur high costs and excessive learning. They are harder to train and require more time. Reducing multidimensional features captures the most relevant features and reduces cost by eliminating noise in the data.

LSA is a NLP method that analyses relationships between a set of documents and the terms contained within. It uses singular value decomposition, a mathematical technique, to scan unstructured data to find hidden relationships between terms and concepts. LSA is an information retrieval technique patented in 1988, although its origin dates back to the 1960s. (Evangelopoulos, 2013)

LSA is primarily used for concept searching and automated document categorization. However, it's also found use in software engineering (to understand source code), publishing (text summarization), search engine optimization, and other applications. (Evangelopoulos, 2013)

There are a number of drawbacks to LSA, the major one being is its inability to capture polysemy (multiple meanings of a word). The vector representation, in this case, ends as an average of all the word's meanings in the corpus. That makes it challenging to compare documents. (Jorge-Botana et al., 2009).

In the world of search engine optimization, Latent Semantic Indexing (LSI) is a term often used in place of LSA. Some marketers believe using LSI can improve on-page SEO. However, given that there are more recent and elegant approaches to NLP, the effectiveness of LSI in optimizing content for search is in doubt (MarketMuse, 2023).

In our project and thesis, LSA is applied after SBERT embeddings as a dimensionality reduction technique. The primary goal of this step is to reduce vector size and eliminate unnecessary noise in the semantic representations.

2.4. Review of Educational Recommendation Applications

In this sub-topic, some papers on the use of recommender systems in education are discussed. The contents of these articles are analyzed. These papers explore various approaches such as collaborative filtering, content-based models, hybrid architectures and chatbots, and neural networks. The selected papers include applications such as personal course recommendations and adaptive educational platforms from learning resource recommendations.

These papers have been selected to address how recommender systems in education have been applied in different settings and how to improve student engagement, performance and personal learning experiences.

Table 1. Overview of Course Recommendation Studies in the Literature

Writer(s)	Year	Objective of the Study	Method	Key Findings
Urdaneta-Ponte et al.	2021	To conduct a comprehensive systematic review of 98 educational recommender system (ERS) studies, focusing on the types of education (formal/non-formal), recommendation algorithms, target elements, and emerging trends. (Urdaneta-Ponte, E., et al., 2021)	Systematic Review	Gaps in personalization and user-centered design
Thongchotchat et al.	2023	To investigate how learning styles are modeled and integrated in ERS research; to classify learning style theories, identification methods, and commonly used recommendation	Literature Review	FSLSM common; hybrid methods proposed

		algorithms. (Thongchotchat, P., et al., 2023)		
Kotsyuba et al.	2022	To design a web-based recommendation service that enables learners to build individualized programming education paths based on skills and labor market alignment. (Kotsyuba, Y., et al., 2022)	Web-Based Recommendation System	Aligns student skills with labor market, reduces decision cost
Aucancela et al.	2023	To examine how various types of ERS impact educational quality globally, using Cochrane methodology to assess effects on graduation rate, performance, and personalization outcomes. (Aucancela, O., et al., 2023)	Systematic Review	Hybrid methods improve graduation and performance metrics
Zeng	2022	To enhance the accuracy and diversity of educational recommendations through a fusion of conditional random fields and graph attention networks in online/offline settings. (Zeng, Y., 2022)	Graph Attention + CRF Model	Enhanced accuracy, adapts to open learning environments
Jin et al.	2021	To develop a personalized mobile recommendation system for language learning apps by leveraging learning styles and AI filtering techniques. (Jin, J., et al., 2021)	Collab. Filtering + FSLSM	High satisfaction, style-preference relationship observed
Chatwattana et al.	2024	To develop a mobile-friendly AI chatbot to	AI Chatbot with SDLC	Very high satisfaction in

		provide personalized educational advice to engineering students, using SDLC for system development. (Chatwattana, P., et al., 2024)		accuracy and usability
Kingchang et al.	2024	To design and evaluate an AI chatbot platform (via LINE app) that offers higher education guidance aligned with user aptitudes. (Kingchang, S., et al., 2024)	AI Chatbot Platform	86% agreement with system suggestions, high usability
Luo et al.	2024	To propose a recommendation approach for MOOCs and study group formation using social network analysis and graph neural networks. (Luo, W., et al., 2024)	GNN + Social Network Analysis	Personalized recommendations and improved collaboration
Pardamean et al.	2022	To predict student learning styles in primary education without questionnaires, and generate personalized recommendations based on AI models. (Pardamean, B., et al., 2022)	AI-based Learning Style Prediction	No questionnaire needed, improved objectivity in suggestions

2.5. Summary of the Literature and Identified Gaps

The importance and functions of recommender systems in education systems are explained. Various studies in this field that address students' personal interests are reviewed.

Despite the growing popularity of recommender systems in education, some needs remain. Most existing systems rely heavily on static student profiles or simple filtering techniques. Few studies rely on rich textual content such as direct course descriptions, learning outcomes or weekly topics.

In addition, there has been limited interest in elective course recommendation systems specifically designed for higher education that integrate students' interests with natural language understanding techniques.

Therefore, this paper aims to address this gap by developing a course recommendation system targeting elective courses in Industrial Engineering at YTU, using SBERT for semantic vectorization and LSA for dimensionality reduction. The proposed system differentiates itself by providing a deeper understanding of students' course interests and providing personalized recommendations.

METHODOLOGY

3.1. System Architecture and Component Overview

The project was developed to enable YTU Industrial Engineering students to choose the right elective course in the field they want to improve themselves.

First, inputs are received from the user regarding the fields they are interested in. "User input is used to rank the courses in the relevant elective category using the dot product.

The conversion of courses into vectors is carried out with high-dimensional semantic vectors using the SBERT model. The data for the model is prepared separately for each course.

In addition, LSA is applied for dimensionality reduction to reduce computational complexity and remove unnecessary features.

An interface has been developed in the project to increase usability and meet user satisfaction. This interface was developed in Django using Python.

In general, the project allows students to explore elective courses, receive personalized recommendations, and review detailed course information.

3.2. Data Collection and Course Attributes

The dataset used in the developed system consists of elective course information from the Bologna Information System for Industrial Engineering at YTU. The recommendation engine focuses on four elective course categories: University Vocational Elective, Social Elective, Vocational Elective A, and Vocational Elective B.

The attributes of each course are as follows:

- <u>Course Name:</u> Contains the name of the course. It serves as the primary identifier but is not used directly for matching.
- <u>Course Code:</u> Each course has its own code. It is used as the primary key in the database.
- <u>Prerequisites:</u> In university curricula, some courses have prerequisites, in other words, courses that must be taken before taking this course. Since this project covers elective courses, this qualification does not mean anything for us.

- Semester: Indicates the semester in which the course is offered. Since it is not clear which elective courses are offered in which semester, this part is not taken into consideration.
- Course language: Indicates in which language the course is offered. These are Turkish or English. Since the language and semester of the courses offered here also change, it is not taken into consideration in order to address the general public. Students who have to take courses in English are assumed to consider the highest rated courses offered in English in the course recommendation system.
- Course objective: This is the part where the basic plan of the course is explained in a few sentences, such as the reason for offering this course, the competencies it aims to provide to the students. This section is used for matching.
- Course Content: Provides a brief overview of the content, objectives and scope of the course. This area is one of the basic texts used for semantic analysis.
- Course Books: It contains the names of the main books used in the course. It has no purpose within our project.
- Weekly Topics: A detailed breakdown of the topics covered each week. The weekly topics are also one of the basic texts used for semantic analysis.
- Learning Outcomes: These are the competencies that students are expected to acquire after completing the course. Learning outcomes are also used in the semantic embedding process to better capture the academic purpose of the course.
- Level (dersin_seviyesi): Contains the value “Bachelor's” for all courses.
- Evaluation System (degerlendirme_sistemi) :The assessment method is determined individually by the course instructor and is not usually specified in the Bologna system, so this column is mostly empty. Therefore, it is not used in the course recommendation process.
- Category (course_category): This column shows the semantic category assigned to each course, such as “Data and Artificial Intelligence” or “Art and Music.” In this project, categories were manually assigned by analyzing the content of each course. These categories play a central role in filtering users' interests and matching them with appropriate courses.

- Presenting Unit (dersi_sunan_birim): Except for university professional elective courses, all elective courses in this project are provided by the Industrial Engineering Department by default.

All text fields (Course Name , Course Category ,Course objective, Course content, weekly topics and learning outcomes) are combined to create a comprehensive "full_text" representation for each course. This combined text is embedded using SBERT to create a semantic vector for each course.

Before embedding, basic preprocessing steps were applied to the merged text. These were removing special characters, deleting unnecessary spaces, adding missing course descriptions, and merging some of the above-mentioned parts that could be sources for the course as course data. Stop word removal or stemming was not applied because the embedding model (SBERT) works on contextual understanding rather than frequency-based word analysis.

The entire course dataset is organized and stored in a structured format using a PostgreSQL database. Each record is uniquely identified by course code, ensuring data integrity and easy access during the recommendation process.

3.3. Embedding Using SBERT and LSA

The course recommendation system uses the SBERT model to generate semantic embeddings for both student input and course descriptions. SBERT is a refined version of the original BERT model, specifically tailored to generate meaningful sentence-level embeddings. Its success at the sentence level is the reason it was chosen for the project.

Reimers and Gurevych (2019) noted that the BERT model, in its direct form, places sentences in a vector space that is not suitable for similarity measurements. To overcome this issue, a model called Sentence-BERT (SBERT) was developed. SBERT adds a pooling operation to BERT's output to produce fixed-sized and semantically meaningful sentence embeddings. Thanks to these vector representations, semantically similar sentences can be found more effectively using measures such as cosine similarity or Euclidean/Manhattan distances (Reimers & Gurevych, 2019).

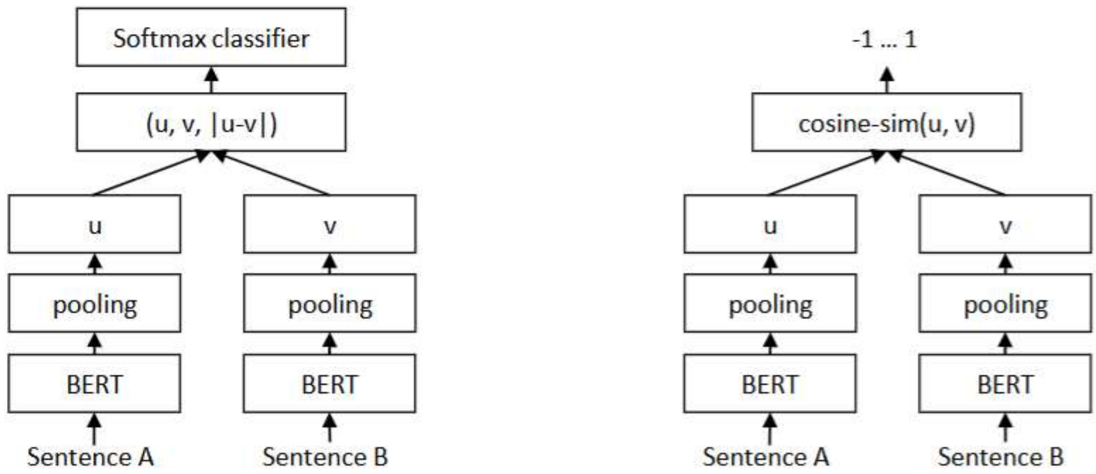

Figure 1. Architecture of SBERT for Sentence Similarity Computation (Reimers & Gurevych, 2019)

The "paraphrase-multilingual-mpnet-base-v2" model was selected for the course recommendation system. This multilingual version of SBERT can effectively encode both Turkish and English texts, which fits well with the bilingual structure of the elective course dataset and the expected user-entered language. Each text is converted into a dense 768-dimensional embedding vector. (Reimers & Gurevych, 2019)

SBERT embeddings tend to be high-dimensional, which causes problems with computational efficiency and potential noise. To address these problems, LSA is applied to the gaps in between.

LSA, implemented via Truncated Singular Value Decomposition (TruncatedSVD), reduces the dimensions of the embeddings while preserving their underlying semantic structure. By projecting high-dimensional embeddings into a lower-dimensional semantic space (typically reduced to 100 components), two important benefits of the system are achieved:
(i) faster similarity programming and
(ii) reduction in noise that would otherwise lead to equality.

After dimensionality reduction with LSA, similarity scores between student input vector and course vectors are calculated using the dot product method. The dot product approach was preferred over cosine similarity to avoid normalization overhead and preserve direct magnitude relationships in the reduced space. (Luo et al., 2017)

The system of embedding with SBERT, dimensionality reduction with LSA, and similarity calculation with dot product provides a balance between semantic accuracy, computational efficiency, and responsiveness, allowing students to receive course recommendations that closely align with their interests.

3.4. Similarity Computation via Dot Product

The dot product is simply the multiplication of the magnitudes and the cosine of the angle between them. Unlike cosine similarity, the magnitudes of the vectors are also important. (Luo et al., 2017)

In this project, dot multiplication is preferred to avoid the normalization overhead required in cosine calculations and to preserve the direct magnitude information embedded in the sentence vectors after dimensionality reduction with LSA. (Luo et al., 2017)

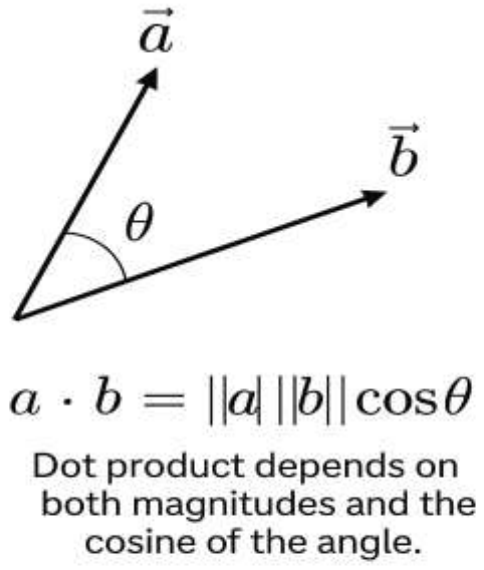

Figure 2. Dot Product Formula and Geometric Interpretation (Luo et al., 2017)

3.5. Web Interface Implementation with Django

The recommendation system interface was developed using the Django framework. It allows YTU Industrial Engineering undergraduate students to explore elective courses, receive personalized recommendations and review detailed course information. The interface architecture is structured in screens that include interest input, course recommendation results, category browsing and individual course details. The user interface allows students to examine the categories as well as the topics under the categories and to analyze the courses that are related to each other. It also enables students who have already taken a course to comment on the course and thus provide guidance for the course.

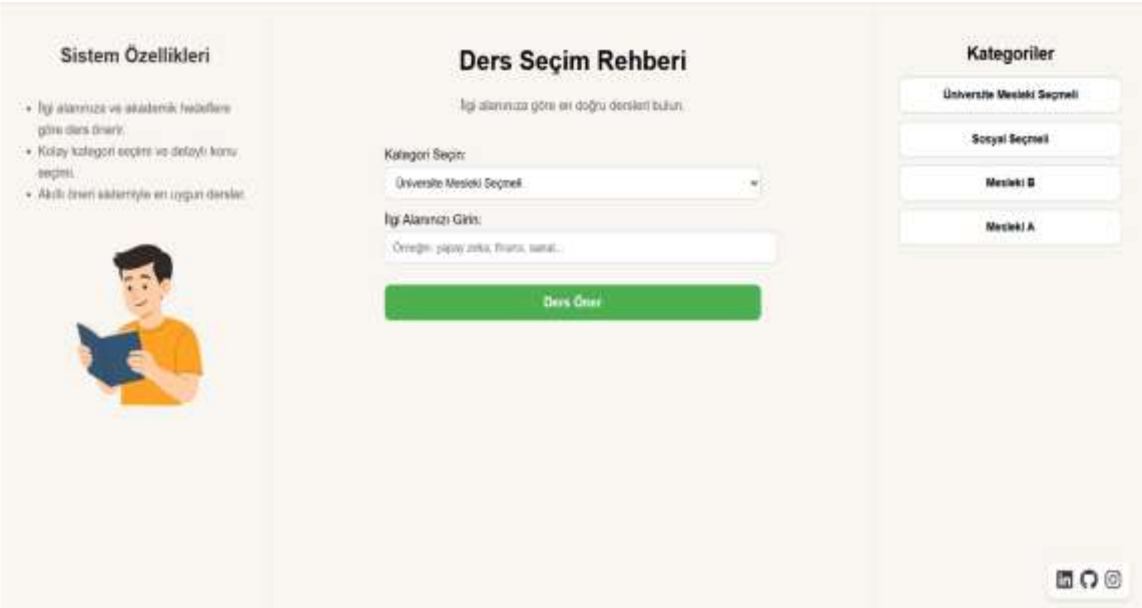

Figure 3. Main Interface of the Course Recommendation System

This screen is the main page of the system. System features are briefly explained in the left panel for students. Categories are shown in the right panel. The middle panel is the most important purpose of the system. Students can access the courses according to the fields they want in the relevant category.

Figure 4. Category Selection Panel of the System

If the student clicks on any of the categories in the right panel, he/she can view the coursesand topics belonging to that category.

Figure 5. Course Recommendation Screen Based on User Interest

In the middle panel, the most meaningful courses are suggested according to the area of interest entered. Students can view the courses they are interested in on this screen and continue to enter other interests.

Figure 6. List of Courses Displayed Within the Selected Category

When students want to see all the courses of a category without any advice, the right panel shows all the relevant courses according to their selection.

Figure 7. Course Detail Page with User Comment Panel

When examining courses in categories, examining courses under the heading of topics or recommended courses, the student can click on the course button and see all the information and content of the relevant course, and if someone has taken the course before, they can also comment on the course.

An example showing the topics.

Figure 8. Topic Headings Displayed by Category Selection

3.6. Definition of the Mathematical Model

In this project, the course recommendation is based on a semantic similarity system. The goal is to find the maximum similarity and the problem is modeled accordingly. The basic assumption is that both the student interest and the course descriptions can be placed in the same high-dimensional semantic space using SBERT. Once embedded, the dot product is used as a similarity measure between the student vector u and each course vector v_i.

The calculation of the S_i is as follows: $S_i = u \cdot v_i$

Table 2. Table of Sets and Parameters

Sets/Parameter	Description
u	embedding vector of the student's input text
v_i	embedding vector of course i
S_i	similarity score for course i

After calculating the Si similarity score for each course, the system ranks all courses in the selected category according to these scores. Courses with the highest similarity score are considered to be the most relevant to the user's interests.

The project is not a maximization problem, but it aims to identify the highest scoring matches using content-based similarity.

RESULTS AND ANALYSES

4.1. Dataset Samples and Category Matching Results

The data set for the project consists of elective courses offered on the YTU Industrial Engineering Bologna page. There are four elective course groups: Professional A, Professional B, University Professional Elective, and Social Elective. Each course is defined by the following 14 descriptive fields:

1. *Course Name (ders_adi)*
2. *Course Code (ders_kodu)*
3. *Prerequisites (on_kosullar)*
4. *Semester Offered (yariyil)*
5. *Language of Instruction (dersin_dili)*
6. *Level (dersin_seviyesi)*
7. *Category (ders_kategorisi)*
8. *Course Objective (dersin_amaci)*
9. *Course Content (ders_icerigi)*
10. *Learning Outcomes (ders_ogrenim_ciktilari)*
11. *Weekly Topics (haftalik_konular)*
12. *Evaluation System (degerlendirme_sistemi)*
13. *Presenting Unit (dersi_sunan_birim)*
14. *Textbooks / Resources (ders_kitaplari)*

	ders_adi character varying (255)	ders_kodu [PK] character	on_kosullar text	yariyil character varying	dersin_dili character varying	dersin_seviyesi character varyin	ders_kategorisi character varying (100)
1	Dijital Pazarlama	MTH3000	Ön Koşul Yok	Güz	Türkçe	Lisans	İşletme ve Finans
2	Yapay Zeka ile Öneri Sistemleri	MTH3005	Ön Koşul Yok	Güz	İngilizce	Lisans	Yapay Zeka ve Veri Bilimi
3	Termodinamik	END2985	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Mühendislik ve Malzeme Bilimi
4	Sistem Dinamiğine Giriş	END3985	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Sistem Analizi ve Dinamikleri
5	Yapay Zeka	END3975	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Yapay Zeka ve Veri Bilimi
6	Malzeme Bilimi	END2975	Ön Koşul Yok	Güz	İngilizce	Lisans	Mühendislik ve Malzeme Bilimi
7	Mukavemet	END2965	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Mühendislik ve Malzeme Bilimi
8	Pazarlama	END3455	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	İşletme ve Finans
9	Uygulamalı Yapay Öğrenme	END3965	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Yapay Zeka ve Veri Bilimi
10	Optimizasyonda Sezgisel Yöntemler	END3875	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Yenilikçi Teknolojiler ve Uygulamalar
11	Finansal Yönetim	END4665	Ön Koşul Yok	Güz	İngilizce	Lisans	İşletme ve Finans
12	İş Hukuku	END4985	Ön Koşul Yok	Bahar	İngilizce, Türkçe	Lisans	İşletme ve Finans
13	Üretim Yöntemleri	END2995	Ön Koşul Yok	Güz	İngilizce, Türkçe	Lisans	Mühendislik ve Malzeme Bilimi
14	Endüstride Bilgisayar Uygulama Yazılım...	END3925	Ön Koşul Yok	Bahar	İngilizce, Türkçe	Lisans	Yenilikçi Teknolojiler ve Uygulamalar
15	Bulanık Kümeler	END3835	Ön Koşul Yok	Güz, Bahar	İngilizce, Türkçe	Lisans	Yenilikçi Teknolojiler ve Uygulamalar

Figure 9. Elective Course Dataset – Part 1

dersi_sunan_birim character varying (100	dersin_amaci text	dersin_icerigi text	ders_kitaplari text	ders_ogrenim_ciktilari text	haftalik_konular text	degerlendirme_sistemi text
Endüstri Mühendis...	Bu ders, dijital pa...	Öğrenciler, dij...	Bilinmiyor	Ders Öğrenim Çıktıları	Hafta 1: (Ön Hazırlık: ...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Bu ders, yapay z...	Öğrenciler, ön...	Bilinmiyor	Ders Öğrenim Çıktıları	Hafta 1: (Ön Hazırlık: ...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Enerji ve dönüşü...	Tanımlar ve T...	Mühendislik Y...	Öğrenciler, termodina...	Hafta 1: Tanımlar ve ...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Öğrencilere siste...	Sistem dinam...	Sterman, J.D. ...	Öğrenci sistem yaklaşı...	Hafta 1: Giriş ve Tem...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Mühendislik uyg...	Yapay zekânı...	C.S. Krishnam...	Öğrenciler yapay zeka...	Hafta 1: Yapay Zekây...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Malzeme yapıları...	Malzeme bili...	Ahmet Topuz, ...	Öğrenciler, üretim sist...	Hafta 1: Malzeme bili...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Mukavemetin te...	Kesit Tesirleri...	Ferdinand P. B...	Öğrenciler kesit tesir d...	Hafta 1: Kesit tesirleri...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Satış yönetimini...	Satıcılık mes...	Güncellenmiş ...	Öğrenciler pazarlama ...	Hafta 1: Satış kavram...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Bu ders, yapay ö...	Öğrenciler, ya...	Bilinmiyor.	Ders Öğrenim Çıktıları	Hafta 1: (Ön Hazırlık: ...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Bu dersin amacı, ...	Bu derste, ön...	Zbigniew Mich...	Öğrenci matematiksel ...	Hafta 1: Giriş: Temel ...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Finansal yönetim...	İşletmelerde ...	Brealey, R. My...	Öğrenciler finansal yö...	Hafta 1: İşletmelerde ...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Öğrencinin çalış...	Hukuk kavra...	Murat DEMİRC...	Çalışma yaşamında ha...	Hafta 1: Hukukun dall...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Teknoloji, üretim ...	Üretim yönte...	Groover, " Mod...	Öğrenciler teknoloji, ür...	Hafta 1: Üretim yönte...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Dersin amacı, öğ...	Bilgisayar Yaz...	Domain Archit...	Öğrenci uygulama yazı...	Hafta 1: Yazılım Kavr...	Devam/Katılım: %; Laborat
Endüstri Mühendis...	Dersin amacı, öğ...	Bulanık mantı...	Fuzzy Logic wi...	Öğrenciler bulanık kü...	Hafta 1: Temel Kavra...	Devam/Katılım: %; Laborat

Figure 10. Elective Course Dataset – Part 2

4.2. Recommendation Outputs from SBERT + LSA

This section shows example outputs of the SBERT+LSA model. Three examples are analyzed by giving different inputs for different categories.

In our first example, for Vocational Elective A, the user has given the input "Makine öğrenmesi ve veri analizi üzerine bir ders istiyorum." As a result, the proposed courses Applied Artificial Intelligence, Recommender Systems with Artificial Intelligence and Artificial Intelligence ranked the most logical answers for Vocational Elective A, which has only 14 courses.

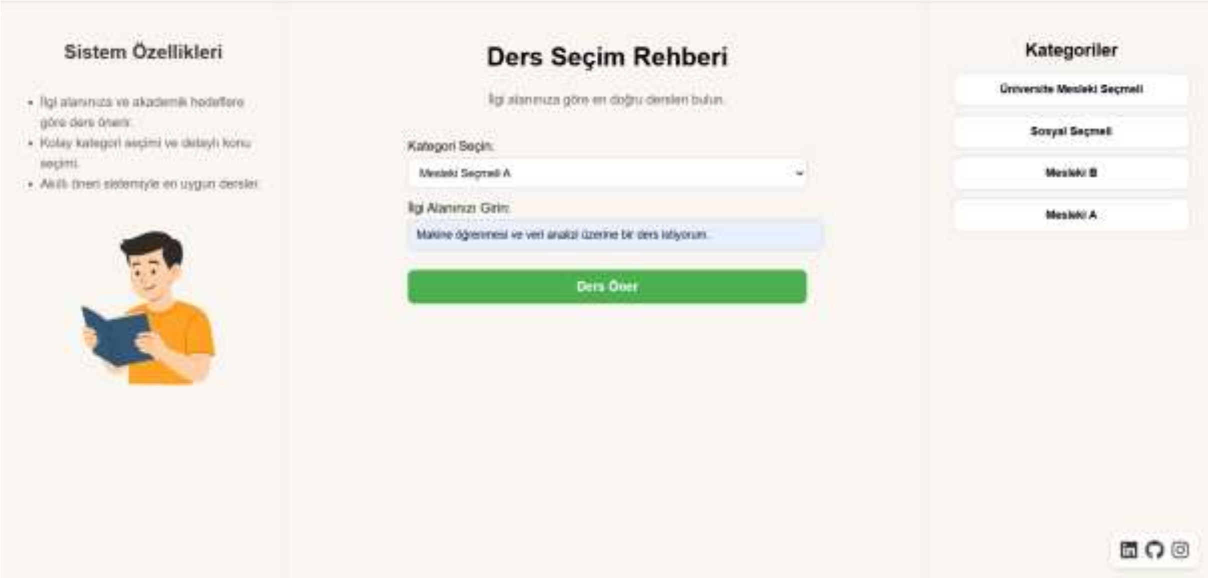

Figure 11. Sample 1 – Input for Course Recommendation

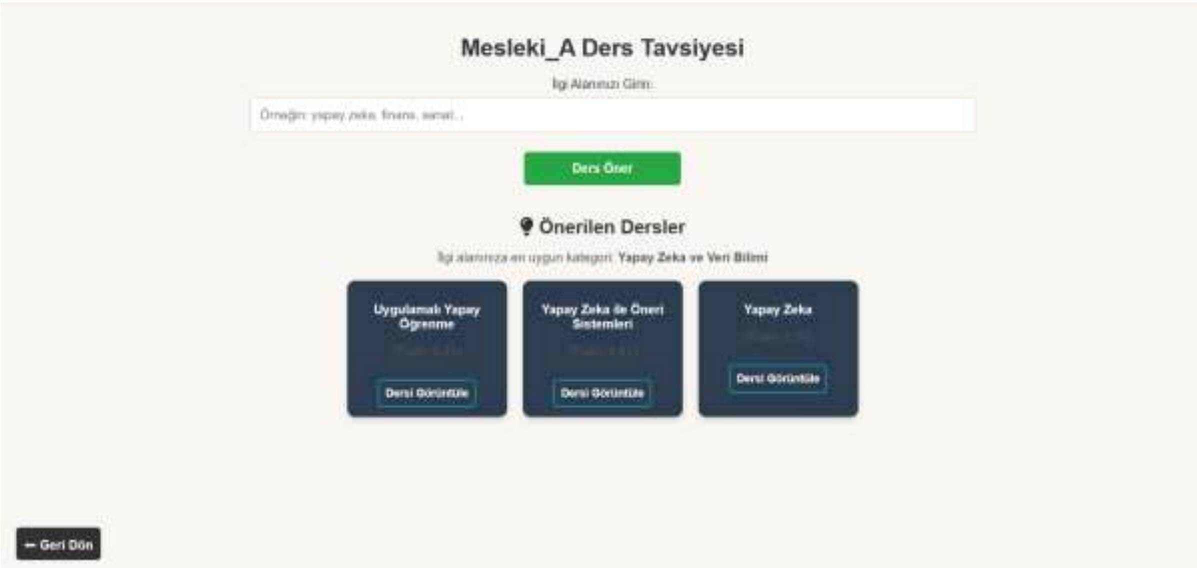

Figure 12. Sample 1 – Output for Course Recommendation

In the second example, the user entered “Yönetim ve liderlik becerilerimi geliştirmek istiyorum.” for the Vocational Elective B category. The suggested courses included Risk Analysis and Management, Strategic Management, Asset Management, Efficiency in Business, and Warehouse Management. These courses overlap with management-oriented courses and demonstrate the success of the SBERT model in distinguishing managerial themes.

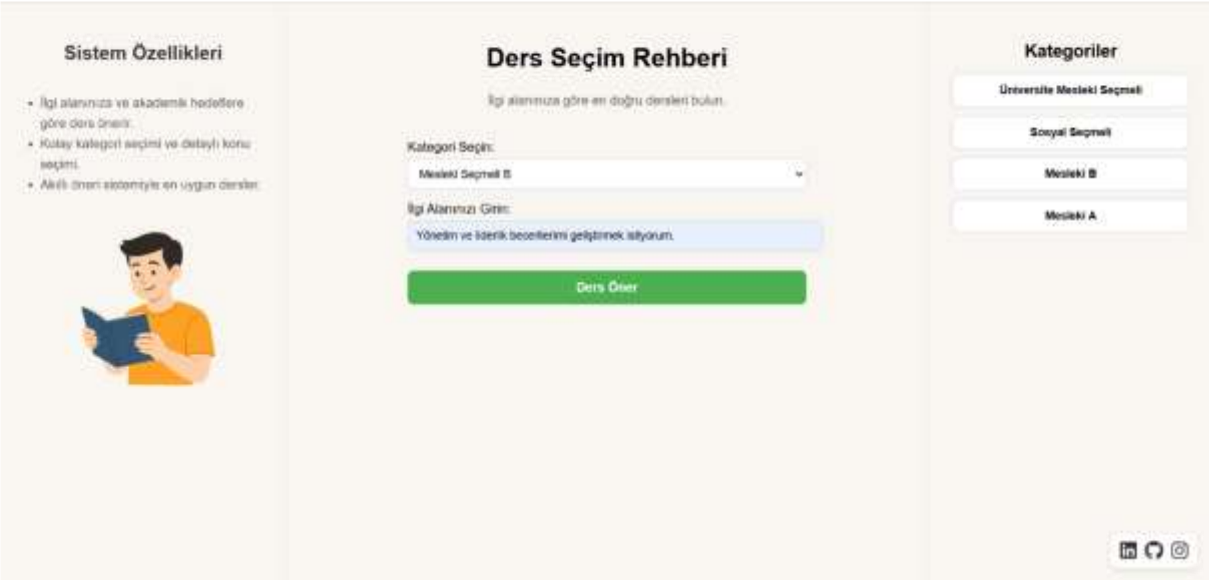

Figure 13. Sample 2 – Input for Course Recommendation

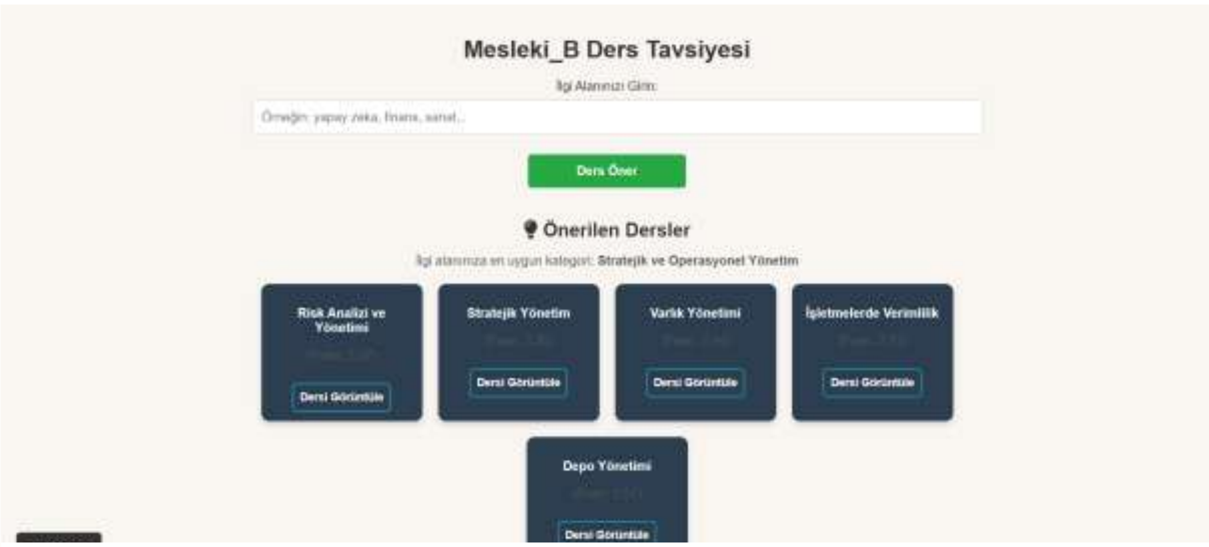

Figure 14. Sample 2 – Output for Course Recommendation

In the third example, the user made an entry in the Social Elective category as "Yabancı dil öğrenmek istiyorum ama latin kökenli dilleri istiyorum." The model suggested Basic Spanish 1, Basic Spanish 2, Basic Italian 1, Basic Italian 2 and Basic French 2. The accuracy of the suggested courses is tested on a specific example, as all the suggested courses cover the user input.

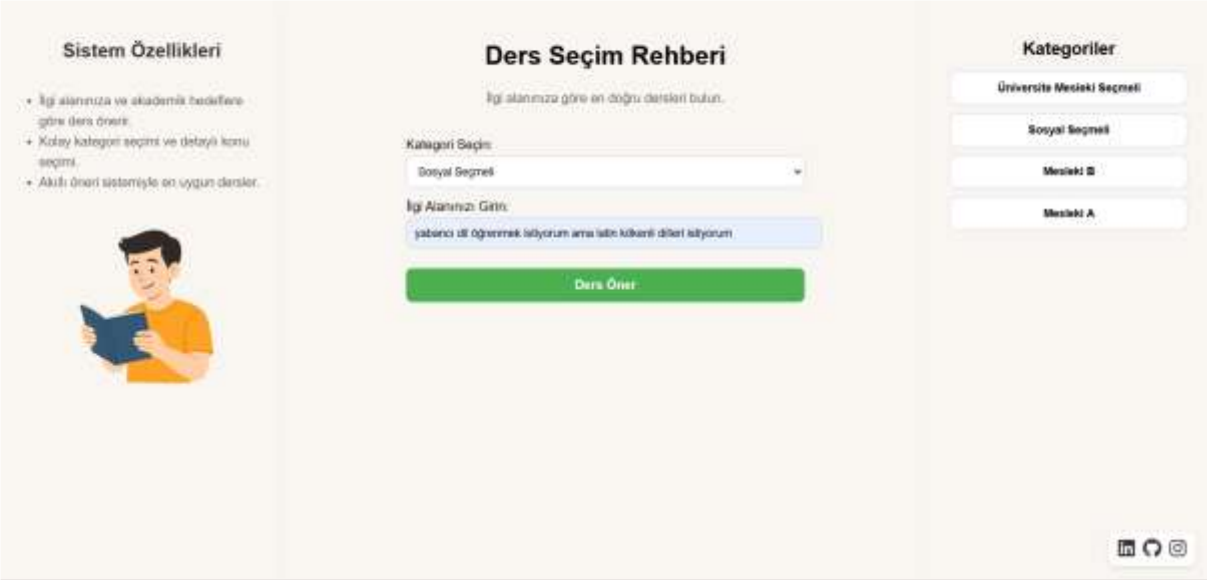

Figure 15. Sample 3 – Input for Course Recommendation

Figure 16. Sample 3 – Output for Course Recommendation

CONCLUSION

5.1. Summary of Findings

In the project, the elective course selection system for Industrial Engineering at Yıldız Technical University was developed using the LSA method for dimension reduction with SBERT-based semantic embeddings. The system aims to provide personalized course recommendations by matching students' interests with course descriptions using dot product similarity.

For example, when the user enters:

“Sanat tarihi ve kültürel değerler hakkında bilgi edinmek istiyorum”

the system automatically determines that the most relevant category is ‘Art and Music.’ Within this category, it presents the most recommended courses along with their similarity scores (dot product values):

- Sanat Tarihi – Score: 5.37
- 16. yüzyıldan Günümüze İstanbul'da Dans – Puan: 3,14

All four elective courses (Vocational Elective A, Vocational Elective B, Social Elective, and University Vocational Elective) in the Industrial Engineering department were included in the project. It successfully generated the best course recommendations based on the queries entered by users. The use of 100-component TruncatedSVD reduced computation time while maintaining semantic quality. Compared to previous keyword-based approaches, this hybrid method demonstrated that it could understand user preferences more accurately, even with short or ambiguous input expressions.

5.2. Contributions and Observations

The project aims to reduce the number of wrong choices made by students transferring to university due to their lack of sufficient knowledge and experience in course selection. In addition, it aims to enable students to gain insight into their desired fields more quickly by taking courses related to those fields. Therefore, it is important that students who write about their areas of interest take the right courses. In particular, the combination of SBERT based embeddings and the LSA method has both increased the accuracy of semantic matching and made the system more efficient through dimension reduction.

Another notable contribution of the project is that it was developed in Turkish. While many recommendation systems in the literature focus on English datasets, this study obtained Turkish data by collecting it individually for each course from YTU's Bologna site. By achieving successful results with Turkish academic content, it has demonstrated that similar systems can be effectively applied in different languages.

Furthermore, rather than any elective course group in Industrial Engineering, all four elective courses in the department were included in the recommendation system.

Observations made during testing revealed that when users entered clear areas of interest (e.g., "Yapay Zeka," "Mühendislik," "Sanat"), the system provided much more successful recommendations. As the user's statement became clearer, it was observed that the model correctly recognized semantic similarities and highlighted appropriate courses. Additionally, it was noted that the quality of recommendations decreased, and similarity scores became closer to each other when entries were made with very general or vague statements. This is a natural and expected outcome.

Furthermore, it is important to benefit from the experiences and opinions of students who have previously taken the course. For this reason, a comment panel has been added for students who have taken each course. This allows new students to gain insight into the course.

Overall, this project combines deep semantic understanding with practical recommendation systems, making it a course selection guide for Industrial Engineering students at Yıldız Technical University.

5.3. Limitations of the Study

This study was conducted within certain limitations. First, the system was developed using only elective course data from the Industrial Engineering Department at YTU. The project does not cover different universities and departments. In addition, course contents were manually obtained from the Bologna system, and some courses had missing information or non-standard expressions. This may affect the model's performance in interpreting some courses.

Another limitation is the direct impact of user input. If users do not clearly express their interests, the quality of the system's recommendations decreases, and it becomes difficult to make a meaningful ranking among courses with similar scores. Furthermore, the courses recommended in the project are based on users' past data.

Finally, the recommendation system is based solely on text similarity; other important variables such as course difficulty level, student success, and instructor are not included in the model. The failure to consider these factors results in users being presented with recommendations based solely on content.

5.4. Suggestions for Future Work

The project will greatly increase the scope of users, especially by expanding it to different universities and departments. Inclusion of departments can be done step by step. In addition, basing the data on past user usage will allow personalized results to be obtained in the recommendations. For example, the recommended courses can not only be based on instant user inputs, but also on individual past data.

In case of reaching a wide user base, the project can be expanded to mobile applications as well as web applications.

REFERENCES

Aucancela, M., González Briones, A., & Chamoso, P. (2023). Educational recommender systems: A systematic literature review. *The Barcelona Conference on Education 2023*.

Chatwattana, P., Yangthisarn, P., & Tabubpha, A. (2024). The educational recommendation system with artificial intelligence chatbot: A case study in Thailand. *International Journal of Engineering Pedagogy, 14*(5), 51–64.

De Boom, C., Van Canneyt, S., Bohez, S., Demeester, T., & Dhoedt, B. (2015). Learning semantic similarity for very short texts. *arXiv preprint arXiv:1512.00765*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(6), 683–692.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM, 35*(12), 61–70.

Jin, L., So, H.-J., Song, J., & Hwang, W. (2021, August). Recommendation system design for AI-integrated educational applications: Focusing on learning styles. In *Proceedings of ICoME 2021* (pp. XX–XX).

Jorge-Botana, G., León, J. A., Olmos, R., & Hassan-Montero, Y. (2009). Visualizing polysemy using LSA and the predication algorithm. *Behavior Research Methods, 41*(3), 712–720.

Kingchang, T., Chatwattana, P., & Wannapiroon, P. (2024). Artificial intelligence chatbot platform for educational recommendations in higher education. *International Journal of Information and Education Technology (IJIET), 14*(1), 34–41.

Kotsyuba, I., Shikov, A., Karpik, K., Galperin, M., Kudriashov, A., & Silko, J. (2022). Recommendation web service for choosing an individual educational path in the field of transportation systems' programming. *Transportation Research Procedia, 63*, 591–599.

Luo, C., Zhan, J., Wang, L., & Yang, Q. (2017). Cosine normalization: Using cosine similarity instead of dot product in neural networks. *arXiv preprint arXiv:1702.05870*.

Luo, Z., Wang, X., Wang, Y., Zhang, H., & Li, Z. (2024). A personalized MOOC learning group and course recommendation method based on graph neural network and social network analysis. *arXiv preprint arXiv:2410.10658*.

Pardamean, B., Suparyanto, T., Cenggoro, T. W., Sudigyo, D., & Anugrahana, A. (2022). AI-based learning style prediction in online learning for primary education. *IEEE Access, 10*, 35725–35735.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 3982–3992.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 175–186.

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data, 9*(1), 1–35.

Thongchotchat, V., Kudo, Y., & Okada, Y. (2023). Educational recommendation system utilizing learning styles: A systematic literature review. *IEEE Access, 11*, 8989–9000.

Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation systems for education: Systematic review. *Electronics, 10*(14), 1611.

Zeng, E. (2022). Application of conditional random field model based on machine learning in online and offline integrated educational resource recommendation. *Mathematical Problems in Engineering, 2022*, Article ID 5746671.