Optimising Student Internet Navigation: A Comparative Analysis of Machine Learning Algorithms for Action Prediction

Omar Zammit Serengul Smith
Clifford De Raffaele

Table of Contents

Introduction
- Problem Definition
- Existing Technologies

Proposed Solution
- Research Methodology
- Data Collection Framework
- Determine Best Algorithm

Implementation

Evaluation

Conclusion

Terminologies

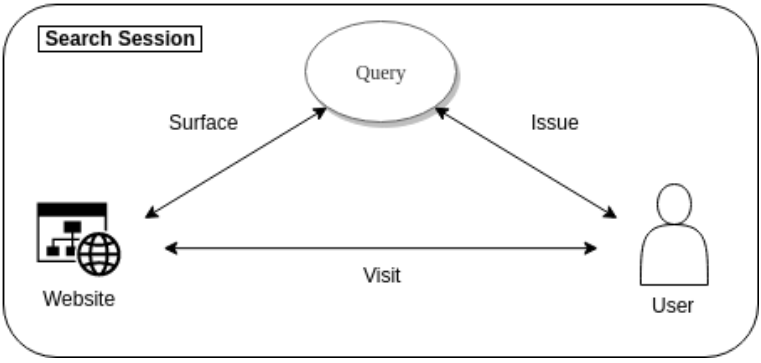


Figure: A web search session as explained in Kim et al. 2012

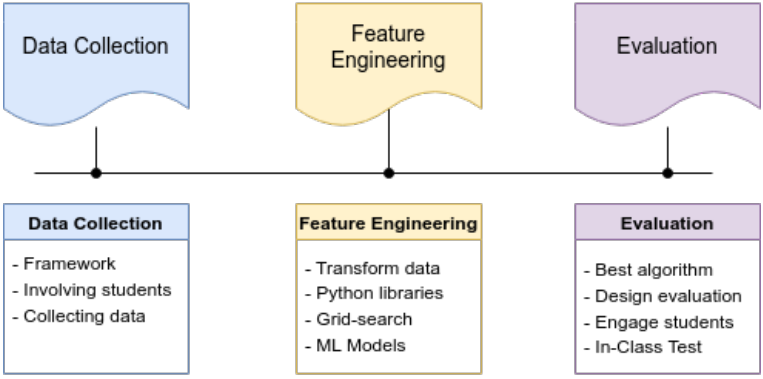
Problem Definition

- Students have access to a wealth of online information and educational resources.
- Students require such information to complete assignments and prepare for exams (Tsai 2009)
- Finding useful information online requires a good searching strategy and prior domain knowledge.
- Novice students who lack domain knowledge will struggle (Debowski 2001).
- Keyphrases searched by students in queries across cohorts are not retained and are forgotten over time.

Next Word Prediction

- An application of Natural Language Processing (NLP).
- Predicting the next most suitable word in a sentence based on preceding words (Rathee and Yede 2023).
- More words more accurate the predictions will be (Rathee and Yede 2023).
- Commonly used in mobile devices
- Improves typing experience (Lehmann et al. 2023)
- **Time-Series** Prediction.
- Forecasting future values and trends (Shi et al. 2023).

Research Methodology



Data Collection Framework

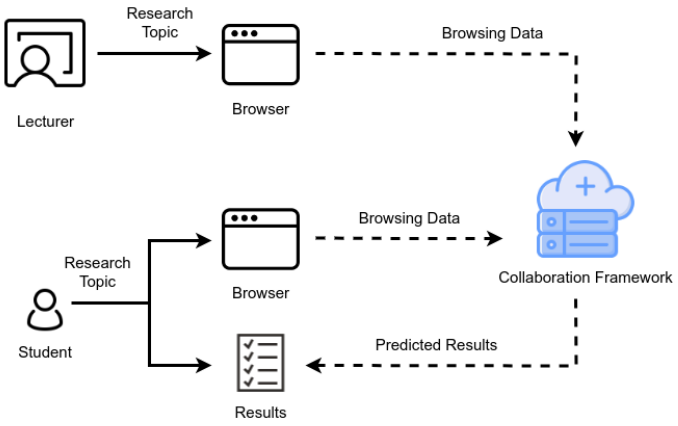


Figure: Collaborative Framework Overview

Collected Data

https://www.google.com/search?q=**machine+learning**

https://stackoverflow.com/search?q=**machine+learning**

https://en.wikipedia.org/wiki/**Machine_learning**

Extraction Keyphrases

Table: Extracting Keyphrases

sequence_key	action_data
22fc4e4e_2023_8_24_18	machine learning
22fc4e4e_2023_8_24_18	artificial intelligence
22fc4e4e_2023_8_24_18	data science
22fc4e4e_2023_8_24_18	supervised learning
22fc4e4e_2023_8_24_18	www.techtarget.com
22fc4e4e_2023_8_24_18	unsupervised learning
22fc4e4e_2023_8_24_18	data science

Constructing Sequences (1 Word)

Table: Sequences (1 Word)

sequence_key	action_data
22fc4e4e_2023_8_24_18	machine learning
22fc4e4e_2023_8_24_18	artificial intelligence
22fc4e4e_2023_8_24_18	data science
22fc4e4e_2023_8_24_18	supervised learning
22fc4e4e_2023_8_24_18	www.techtarget.com
22fc4e4e_2023_8_24_18	unsupervised learning
22fc4e4e_2023_8_24_18	data science

Constructing Sequences (3 Word)

Table: Sequences (3 Word)

sequence_key	action_data
22fc4e4e_2023_8_24_18	machine learning
22fc4e4e_2023_8_24_18	artificial intelligence
22fc4e4e_2023_8_24_18	data science
22fc4e4e_2023_8_24_18	supervised learning
22fc4e4e_2023_8_24_18	www.techtarget.com
22fc4e4e_2023_8_24_18	unsupervised learning
22fc4e4e_2023_8_24_18	data science

Metrics Used for Evaluation

Precision:
Precision measures the accuracy of positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (Sensitivity):
Recall measures the ability of the model to capture all positive instances.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 Score:
The F1 score is the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results

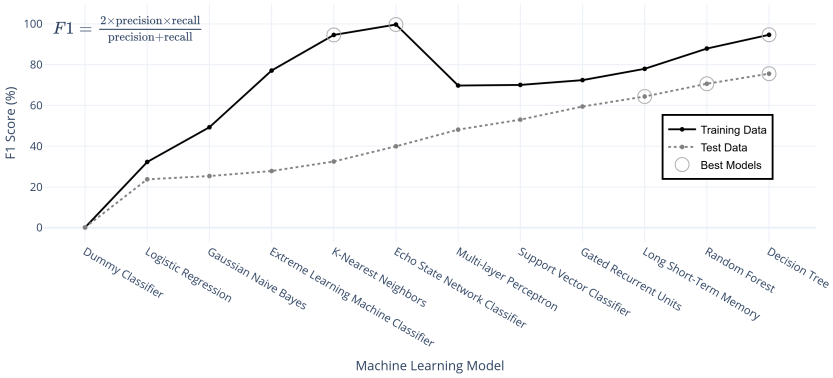
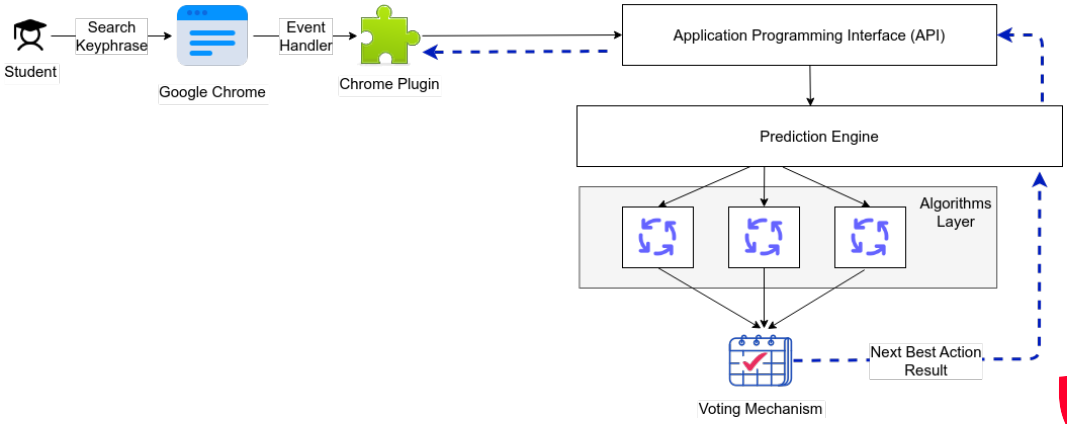
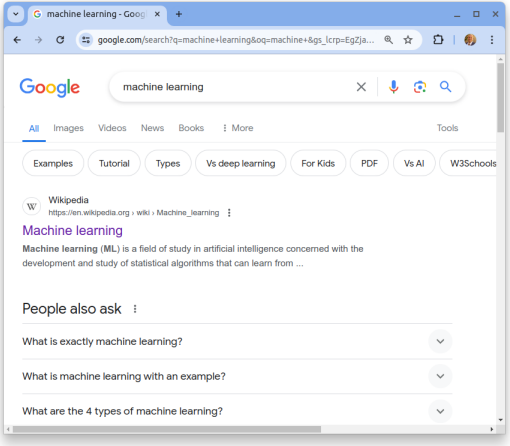


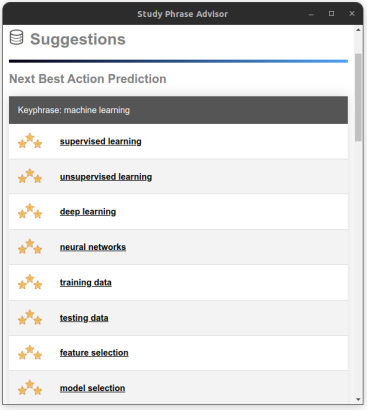
Figure: F1 scores obtained after the grid search.

Prediction Engine

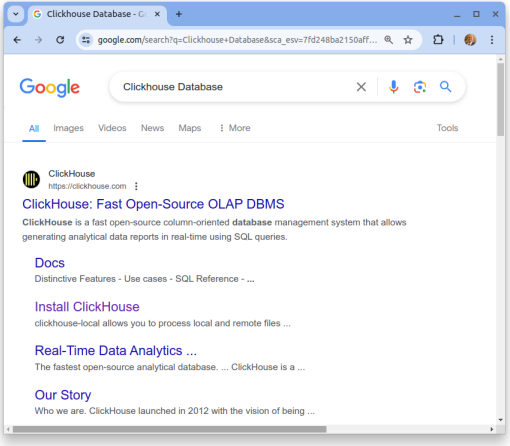


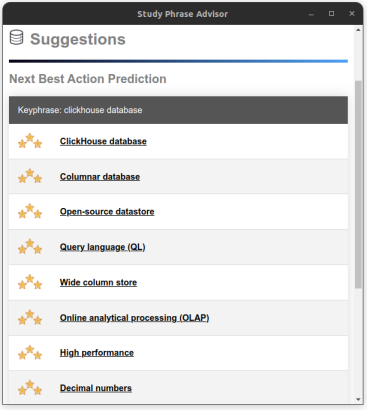
Example 1: Training During a Machine Learning Class



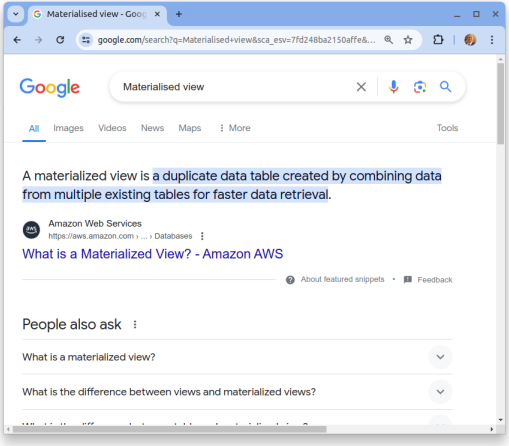


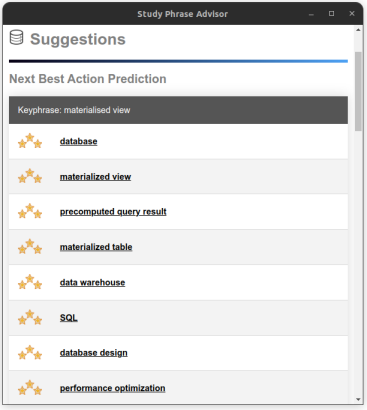
Example 2: Training While Preparing For a Lecture



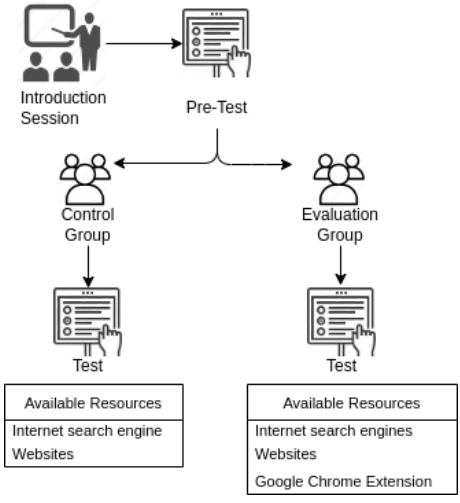


Example 3: Lacking When Searching Some Keyphrases





In-Class Test design



- Brainstorming session with the subject matter lecturer to design tests.
- Focused on topics not yet discussed during the lectures.
- Closed-ended multiple choice was used to avoid subjective answers.
- Question type included: Acronymns, Problem Solving and Theoretical.
- 6 students acted as the control group while 5 as the evaluation group.

Aggregated Results

Table: Evaluation Results

	Evaluation	Control
Total Students	5	6
Pre-Test Average Score ($\bar{P}$)	2.00%	6.67%
Test Average Score ($\bar{T}$)	84.00%	58.33%
Difference ($\bar{P} - \bar{T}$)	82.00%	51.66%
Std Dev (σ)	19.24	19.41

Pros and Cons

- ✓ Knowledge gain - The evaluation group performed better.
- ✓ Motivation - Students showed a high level of enthusiasm while using the application.
- ✓ Novice Students - Such students will gain advantages from key phrases extracted from modules of previous years.
- X Privacy - Students expressed concerns regarding the data collection system during browsing.
- X Unseen Data - When the system lacks training on a particular topic, the suggestions provided may be ambiguous.

Conclusion

- Data collection, model selection and evaluation of various machine learning algorithms.
- Results showed that the evaluation group performed better.
- The approach provided a pedagogical benefit.
- Future research opportunities include evaluation across diverse educational domains and student populations.

Thank You

Omar Zammit ozammit@ieee.org
Serengul Smith s.smith@mdx.ac.uk
Clifford De Raffaele c.deraffaele@mdx.ac.uk