EAI Endorsed Transactions
on e-Learning
Research Article EAI.EU

Ethic wars: student and educator attitudes in the context of ChatGPT

Süleyman Eken[1,*]

[1]Kocaeli University, Department of Information Systems Engineering, Izmit 41001, Turkey

Abstract

Technologists and educators have been both fascinated and frightened since the publication of ChatGPT. ChatGPT has both supporters and detractors, but it is informative for individuals in the education community to look at the educational research on AI in education in order to gain understanding and establish objective judgments about the importance of ChatGPT in education. In this paper, we first present the journey of OpenAI GPT models, then give the implications of ChatGPT for education. Then, we list works for detection ChatGPT based texts and other precautions. Finally, an example of an exam with ChatGPT answers is given.

Received on 21 January 2024; accepted on 29 January 2024; published on 29 January 2024

Keywords: ChatGPT, Education, Ethics, Large language models

doi:10.4108/eetel.4917

1. Introduction

Online learning has traditionally been seen as an option that could benefit a particular demographic of students who are noticeably older and have more familial, financial, and employment-related commitments. The COVID-19 pandemic [1–3], which caught everyone off guard in December 2019, however, drastically altered the educational landscape in a matter of weeks [4]. In order to immediately respond to the challenges provided by the pandemic, several higher education institutions have switched to online coursework, testing, and project. Particularly with regard to the switch to online exams, there have been worries about the likelihood of plagiarism and other types of academic dishonesty [5]. Because students could have access to and exchange resources during the exams. Academic integrity has been protected in online exams using a variety of technical tactics, including proctored exams, plagiarism detection software, and exam security measures.

The academic integrity of online tests, particularly those involving high-order reasoning, is now under threat. The recent public release of Chat Generative Pre-Trained Transformer (ChatGPT) by OpenAI [6] has significantly improved the world's understanding of AI's capacity for natural language processing and (NLP) and natural language understanding (NLU) in a wide range of applications, industries, and job roles. Here are a few illustrations: (i) Consumer service: ChatGPT can enable virtual assistants and chatbots that respond to customer questions and offer useful information. (ii) ChatGPT can be used to create written material, including emails, articles [7**?**], and product descriptions. (iii) Language Translation: ChatGPT can power machine translation systems, simplifying communication with speakers of other languages. (iv) Data Analysis: ChatGPT can be used to analyze vast quantities of text data and derive insights that can be applied to fields like marketing, finance, and healthcare. (v) Search Engine: By doing this, users can enter inquiries without having to use specific words or phrases, just like they would when asking a person a question. For many people, this can make the search process more simple and user-friendly. (vi) Coding: This enables developers to enter code fragments or precise commands just like they would when speaking to a person. Also, ChatGPT is an effective tool for code generation because it can also produce new code [8].

The development of ChatGPT and other cutting-edge language processing tools may result in automation and improved productivity in some sectors and employment categories. Jobs that require repeatable duties, like data input and customer service, may fall

*Corresponding author. Email: suleyman.eken@kocaeli.edu.tr

under this category. It's also vital to keep in mind that the use of such technologies might potentially open up new career prospects in fields like management, programming, and data analysis. A human-in-the-loop (HIL) strategy is also employed in many industries where chatbots or language models like ChatGPT can support people but not entirely replace them.

However, there are issues with cybersecurity and privacy, just like with any technology that can produce text that seems like it was written by a human. Deepfake text has the potential to be one of the key hazards. This can be used to propagate false information or impersonate someone online. Since the model was pre-trained using a sizable dataset of online content, there is also the potential of sensitive data or biases being leaked. If the model is applied to data containing sensitive information, such as personal, financial, or health information, privacy violations may result [9].

2. The Journey of OpenAI GPT Models

By offering extremely potent language models, OpenAI's Generative Pre-trained Transformer (GPT) models have taken the NLP world by storm. Without any supervised training, these models are capable of carrying out a variety of NLP tasks like question answering, textual entailment, text summarization, etc. These language models perform equally well as or even better than the most advanced models trained in a supervised manner and require extremely few to no examples to comprehend the tasks. We'll discuss these models' journey in this part and see how they changed over the course of two years.

Prior to GPT-1 [10], the majority of cutting-edge NLP models were supervised learning trained particularly for a given goal, such as sentiment categorization or textual entailment. Nevertheless, supervised models have two significant drawbacks: (i) They require a significant amount of labeled data, which is frequently difficult to find, in order to master a specific task. (ii) They are unable to generalize to tasks for which they were not specifically taught. Radford et al. suggested building a generative language model from raw data and then refining it using examples of certain downstream tasks like textual entailment, sentiment analysis, and classification. The language model for GPT-1 was trained using the BooksCorpus dataset. The approximately 7000 unpublished books in BooksCorpus (about 5 GB) helped train the language model on omitted data. To train a language model, GPT-1 uses a 12-layer decoder-only transformer structure with masked self-attention. The model had 117M parameters in total.

The improvements to the GPT-2 model [11] mainly involved utilizing a larger dataset and more parameters to the model in order to create an even more powerful language model. The authors scraped the Reddit platform and extracted data from outbound links of highly upvoted posts in order to produce a large and high-quality dataset. The end output, dubbed WebText, had 40GB of text information from more than 8 million papers. With 1.5 billion parameters, GPT-2 had ten times as many as GPT-1. In zero-shot settings, GPT-2 improved former state-of-the-art for 7 of the 8 language modeling datasets.

Open AI created the GPT-3 model [12] with 175 billion parameters in its effort to create extremely robust and potent language models that would require little training and only a few demos to comprehend tasks and carry them out. This model featured 100 times more parameters than GPT-2 and ten times more than Microsoft's potent Turing NLG language model. GPT-3 performs well on downstream NLP tasks in zero-shot and few-shot settings because of the numerous parameters and sizable dataset it was trained on. Five distinct corpora were used to train the GPT-3, each with a specific weight. These include Wikipedia, WebText2, Books1, Books2, and Common Crawl.

Built on top of the GPT-3 family of big language models from OpenAI, ChatGPT (GPT-3.5) [13] is customized using supervised and reinforcement learning methods. ChatGPT only knows items it learned before 2021, unlike search engines (like Google, Bing, or Baidu), which crawl the web for information on current events. The model has more than 117M parameters in total.

3. Implications of ChatGPT for Education

Students: Some advantages are: (i) Interactive and Engaging: ChatGPT offers students a fun and interactive method to study and comprehend difficult ideas. They can interact with the model in real-time, ask questions, and get prompt answers. (ii) Personalized Learning: ChatGPT can offer students individualized learning experiences by adjusting to their level of comprehension and delivering content that is suited to their particular needs. (iii) Access to Information: ChatGPT offers access to a wealth of expertise and information, making it a useful tool for students looking for clarification on their issues. (iv) Cost-Effective: Using ChatGPT as a learning tool is less expensive than traditional classroom instruction because it does not require teachers or physical classrooms. (v) Enables Self-directed Learning: ChatGPT encourages students to take charge of their own learning and become more self-directed by giving them access to information and immediate feedback. (vi) Enhances Critical Thinking: As students interact with ChatGPT, they are required to use critical thinking, problem-solving, and analysis, which can enhance their overall critical thinking abilities [14, 15].

Some disadvantages are: (i) Lack of Human Interaction: Although ChatGPT gives students easy access to material, it lacks the emotional connection and one-on-one communication that they can have with a real teacher. (ii) Limited Feedback: ChatGPT can offer immediate feedback, but it might not always be thorough or nuanced enough to properly address a student's queries or concerns. (iii) Potential for Misinformation: Since ChatGPT was trained on a sizable corpus of text, it's possible that the information it provides is inaccurate or out-of-date. Students should exercise critical thinking and double-check ChatGPT's information. (iv) Dependence on Technology: Because ChatGPT relies on reliable technology and internet connectivity, it may be difficult for certain students to use it as a learning tool if they lack those tools. (v) Lack of Innovative Teaching Techniques: The main structure of ChatGPT is a question-and-answer one, which may not always be the most interesting or useful for teaching. (vi) Limited Capability to Provide Hands-On Experience: Because ChatGPT is mostly text-based, it might not be able to give students opportunities to apply principles practically. (vii) Risk of Overreliance: Although ChatGPT can be a useful tool for students, they should not rely solely on it for their education and should interact with other sources and materials as well in order to have a well-rounded education.

Educators: With its capacity to automate tedious processes and offer individualized guidance to pupils, ChatGPT has the potential to revolutionize education. A few instances of how it might be applied in the classroom are as follows: (i) Grading made simple: Visualize having a program that could evaluate essays and comments from students, giving you more time to give helpful criticism and support. (ii) Personalized input: Using ChatGPT, teachers can create feedback that is specifically suited to each student, assisting them in better understanding and performance. (iii) Dynamic materials: Using ChatGPT to produce questions, quizzes, and practice problems based on certain themes or learning objectives, you may rapidly and effectively create interesting and pertinent instructional resources. (v) Comprehensive lesson planning: AI systems can provide teachers with access to a wealth of knowledge and resources, allowing them to develop more precise and thorough lesson plans. ChatGPT is a really useful place to start when coming up with materials, lesson plans, and plan summaries. AI has a lot to offer in terms of time savings for lesson planning and resource preparation. Here are some instances of how ChatGPT can help you finish lesson planning and resource preparation faster: quick question generation, gap-fill activities, design math and science word problems, writing examples, writing feedback, personalization and differentiation, discussion prompts, one-on-one tutoring or coaching, and letters and communications [16–20].

On the other hand, there are some disadvantages. (i) Materials produced by AI might be prejudiced or contain inaccuracies, particularly if they are not thoroughly evaluated by humans before usage. (ii) Lessons that are overly reliant on the instrument are likely to be highly standardized. Planning and preparing for the learning demands in our classrooms is still of utmost importance. (iii) Some teachers may find it difficult or expensive to use AI tools in the future. The creativity and human touch that go into producing top-notch lesson plans and instructional resources cannot be entirely replicated by AI systems.

4. Works for Detection ChatGPT-based Texts and Other Precautions

Researchers explore the zero-shot machine-generated text identification problem, in which they employ various probabilities computed by a generative model to ascertain whether a candidate passage was sampled from it. This topic has high stakes and frequently sees the creation of new large language models (LLMs). A new curvature-based criterion for determining if a passage is produced from a certain LLM is defined by Mitchell et al [21]. The Human ChatGPT Comparison Corpus (HC3) is a dataset collected by Gu et al [22]. They investigate the properties of ChatGPT's responses, as well as the discrepancies and gaps from a human expert, using the HC3 dataset.

Other precautions can be listed as follows: OpenAI is developing a tool to try to watermark its text creation systems in an effort to fight bad actors utilizing their services for academic plagiarism or for spam, according to OpenAI guest researcher Scott Aaronson [23, 24]. Due to the factually ambiguous nature of ChatGPT's responses, the question-and-answer website Stack Overflow banned the use of ChatGPT in December 2022 [25]. The International Conference on Machine Learning outlawed using ChatGPT or LLMs to create any text in submitted papers without proper documentation in January 2023 [26]. A tool called "GPTZero," developed by Edward Tian [27], a senior undergraduate student at Princeton University, can be used to evaluate how much of a text is AI-generated, making it useful for preventing academic plagiarism. Another discussion is about whether ChatGPT should take place as a co-author (artificial author) in academic studies [28].

Even though it has been stated that methods to detect ChatGPT use have been developed, it is anticipated that solutions would be developed to get around such detectors. Instead of using such techniques, adopt ones that promote learning and discourage usage. An important strategy for preventing the use of ChatGPT

or other analogous technologies is to create assignments that demand students demonstrate progress on their work. These strategies increase the likelihood that the end product reflects the students' efforts rather of being merely a copy of the work of others, encourage deeper learning, and provide teachers more regular access to their students' work (human or artificial). Writing tasks should only be handwritten in class to avoid plagiarism associated to ChatGPT, according to recent articles. While this would surely address the issue of students using ChatGPT, some students may find these exercises to be very difficult, such as those who need more time to finish assignments and those whose first language is not English and who need additional time and resources to write [29].

5. An Example of an Exam with ChatGPT Response: Introduction to Big Data Analysis

In this section, ChatGPT's answers to the online exam of the "Introduction to Big Data Analysis" course given in Kocaeli University Information Systems Engineering Department in 2020 will be investigated [30]. Table 1 shows the content of each section of questions and the ability to be solved by ChatGPT. All questions are given in Appendix. General evaluations and sample outputs are as follows.

Section 1 is related to math and probability questions. ChatGPT could not respond to these ones. But MathGPT [31] could do.

Section 2 includes machine learning model evaluation part. ChatGPT could not respond to the 2.3. Instead of it, a Python-based code is presented. As an example, response to the 2.1:

> Underfitting occurs when a model is too simple and doesn't capture the complexity of the underlying relationship between the features and the target variable. In other words, the model has a poor fit to the training data and fails to generalize to new unseen data. This is shown in the first figure, where the model is represented by a straight line that does not follow the pattern of the data points.
>
> Overfitting, on the other hand, occurs when a model is too complex and fits the noise in the training data, rather than the underlying relationship. The model becomes too specific to the training data and fails to generalize well to new unseen data. This is shown in the second figure, where the model is represented by a highly curved line that fits the noise in the training data, but would not accurately predict the target variable for new data.
>
> Both underfitting and overfitting can result in poor model performance, and finding the right balance between model complexity and fit to the data is key to building a good model.

Section 3 includes dimensionality and feature selection process. Figure interpretation-based can not be responded. As an example, response to the 3.3:

> A dataset is considered balanced if it has a roughly equal number of samples for each class. In a binary classification problem, for example, a balanced dataset would have roughly 50% of samples labeled as positive and 50% labeled as negative. In multiclass problems, a balanced dataset would have a roughly equal number of samples for each class.

Section 4 big data analytics overview. ChatGPT obeys the limitations within the question such as "in a max of two sentences". Response to the 4.1:

> The analytical hierarchy is a systematic way of approaching complex decisions by breaking them down into smaller, simpler parts. The knowledge hierarchy organizes information into categories of increasing specificity, while the value hierarchy ranks values and goals based on personal importance.

Section 5 includes Haddop design patterns. Response to the 5.1:

> Map Function:
>
> Input: (key, value) where key is the rectangle ID and value is a string of rectangle information, including the x and y coordinates.
>
> Output: (key, value) where key is a pair of rectangle IDs and value is a string indicating whether the rectangles intersect.
>
> 1.Split the value string into individual components (rectangle ID, x1, y1, x2, y2)
>
> 2.For each rectangle, generate a list of rectangle ID pairs along with a value indicating whether the rectangles intersect
>
> 3.Emit the rectangle ID pairs as the key and the intersection status as the value
>
> Reduce Function:
>
> Input: (key, values) where key is a pair of rectangle IDs and values are the intersection status.

Table 1. ChatGPT responses to exam sections

Section #	Content	ChatGPT
1	math and probability	not solved
2	book knowledge and pseudo-code	partially solved
3	figure interpretation and book knowledge	partially solved
4	book knowledge but with limitations	solved
5	book knowledge and design pattern	solved
6	real-life design project	solved

> Output: (key, value) where key is a pair of rectangle IDs and value is the final intersection status.
>
> 1.If any of the values for a given key is "intersect", emit the key as the key and "intersect" as the value
>
> 2.If all of the values for a given key are "not intersect", emit the key as the key and "not intersect" as the value
>
> This MapReduce algorithm will find intersecting rectangles by mapping rectangle information to rectangle pairs and checking if they intersect. The intersecting rectangles are then reduced to a final list of intersecting rectangle pairs.

Section 6 includes designing a real-life project. Response to the 6:

> Problem Statement: Analyze customer purchase behavior to find patterns and make recommendations to increase sales.
>
> Solution:
>
> 1.Collect and clean customer purchase data from multiple sources (e.g. transactional databases, log files).
>
> 2.Load the cleaned data into a Spark DataFrame and use Spark SQL to run exploratory analysis and identify patterns.
>
> 3.Use Spark MLlib to build machine learning models to predict future customer purchases based on their past behavior.
>
> 4.Use Spark Streaming to process real-time customer purchase data, feed the data into the machine learning models, and make real-time recommendations to customers.
>
> 5.Visualize the results using tools such as Databricks, Tableau, or Matplotlib to share insights and help drive business decisions.

As seen in ChatGPT responses, the most of questions are solved by ChatGPT and it passed the exam. Also, please note that ChatGPT has passed several exams such as law and medical exams with the help of humans [32].

6. Conclusion

ChatGPT is an AI language model; it does not possess a personality or feelings. However, its educational powers and constraints can be discussed objectively. With its ability to produce text responses that resemble those of humans and its training on a sizable dataset, ChatGPT can be used to generate content and respond to queries on a variety of subjects, including education. It can give information, clarifications, and summaries of educational topics, although it might not always give the most precise or recent data. Furthermore, ChatGPT is unable to comprehend the context and linguistic nuance in the same way that a human does, which could lead to simplistic or inaccurate responses, particularly for complicated and ambiguous topics. In conclusion, ChatGPT can be a helpful tool for creating educational content and providing answers, but it shouldn't be relied on exclusively for learning because it can't comprehend and interpret complicated educational concepts in a way that is human-like.

References

[1] Yurtsever, M.M.E., Shiraz, M., Ekinci, E. and Eken, S. (2023) Comparing covid-19 vaccine passports attitudes across countries by analysing reddit comments. *Journal of Information Science* : 01655515221148356.

[2] Özgüven, Y.M. and Eken, S. (2023) Distributed messaging and light streaming system for combating pandemics: A case study on spatial analysis of covid-19 geotagged twitter dataset. *Journal of Ambient Intelligence and Humanized Computing* **14**(2): 773–787.

[3] Eken, S. (2023) A topic-based hierarchical publish/subscribe messaging middleware for covid-19 detection in x-ray image and its metadata. *Soft Computing* **27**(5): 2645–2655.

[4] Hussein, E., Daoud, S., Alrabaiah, H. and Badawi, R. (2020) Exploring undergraduate students' attitudes towards emergency online learning during covid-19: A case from the uae. *Children and youth services review* **119**: 105699.

[5] Susnjak, T. (2022) Chatgpt: The end of online exam integrity? *arXiv preprint arXiv:2212.09292* .
[6] (2023), Openai, https://openai.com/.
[7] O'Connor, S. *et al.* (2022) Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice* **66**: 103537–103537.
[8] Gozalo-Brizuela, R. and Garrido-Merchan, E.C. (2023) Chatgpt is not all you need. a state of the art review of large generative ai models. *arXiv preprint arXiv:2301.04655* .
[9] Mijwil, M., Aljanabi, M. *et al.* (2023) Towards artificial intelligence-based cybersecurity: the practices and chatgpt generated ways to combat cybercrime. *Iraqi Journal For Computer Science and Mathematics* **4**(1): 65–70.
[10] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. *et al.* (2018) Improving language understanding by generative pre-training .
[11] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. *et al.* (2019) Language models are unsupervised multitask learners. *OpenAI blog* **1**(8): 9.
[12] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A. *et al.* (2020) Language models are few-shot learners. *Advances in neural information processing systems* **33**: 1877–1901.
[13] Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., Uribe, J.F.C. *et al.* (2022) Chatgpt: Optimizing language models for dialogue. *OpenAI blog* .
[14] Rudolph, J., Tan, S. and Tan, S. (2023) Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching* **6**(1).
[15] Altarawneh, H. (2023) Chatgpt impact on student educational performance: a conceptual analysis. *EAI Endorsed Transactions on e-Learning* **9**.
[16] (2023), What is chatgpt and what are the implications for education?, https://www.prosperoteaching.com/blog/2023/01/what-is-chatgpt-and-what-are-the-implications-for-education.
[17] Lo, C.K. (2023) What is the impact of chatgpt on education? a rapid review of the literature. *Education Sciences* **13**(4): 410.
[18] Adeshola, I. and Adepoju, A.P. (2023) The opportunities and challenges of chatgpt in education. *Interactive Learning Environments* : 1–14.
[19] Grassini, S. (2023) Shaping the future of education: exploring the potential and consequences of ai and chatgpt in educational settings. *Education Sciences* **13**(7): 692.
[20] Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J.M. and López-Meneses, E. (2023) Impact of the implementation of chatgpt in education: A systematic review. *Computers* **12**(8): 153.
[21] Mitchell, E., Lee, Y., Khazatsky, A., Manning, C.D. and Finn, C. (2023) Detectgpt: Zero-shot machine-generated text detection using probability curvature. *arXiv preprint arXiv:2301.11305* .
[22] Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J. *et al.* (2023) How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597* .
[23] Kovanovic, V. (2022), The dawn of ai has come, and its implications for education couldn't be more significan?, https://theconversation.com/the-dawn-of-ai-has-come-and-its-implications-for-education-couldnt-be-more-significant-196383.
[24] KWiggers, K. (2022), Openai's attempts to watermark ai text hit limits, https://theconversation.com/the-dawn-of-ai-has-come-and-its-implications-for-education-couldnt-be-more-significant-196383.
[25] Vincent, J. (2022), Ai-generated answers temporarily banned on coding q&a site stack overflow, https://www.theverge.com/2022/12/5/23493932/chatgpt-ai-generated-answers-temporarily-banned-stack-overflow-llms-dangers.
[26] Vincent, J. (2023), Top ai conference bans use of chatgpt and ai language tools to write academic papers, https://www.theverge.com/2023/1/5/23540291/chatgpt-ai-writing-tool-banned-writing-academic-icml-paper.
[27] Tian, E. (2023), Gptzero, https://gptzero.me/.
[28] Stokel-Walker, C. (2023) Chatgpt listed as author on research papers: many scientists disapprove. *Nature* **613**(7945): 620–621.
[29] Cotton, D.R., Cotton, P.A. and Shipway, J.R. (2023) Chatting and cheating: Ensuring academic integrity in the era of chatgpt. *Innovations in Education and Teaching International* : 1–12.
[30] Eken, S. (2020) An exploratory teaching program in big data analysis for undergraduate students. *Journal of Ambient Intelligence and Humanized Computing* **11**(10): 4285–4304.
[31] (2023), Mathgpt, https://mathgpt.streamlit.app/.
[32] Mohmad, P., Ai bot chatgpt passes law and medical exams with human help.

Appendix

Introduction to big data analysis – Midterm exam

1. (Lineer Algebra, Statistics & Probability Fundementals)

1.1.*(6p)* Let the matrix M and v be given. What would be the output of the following code snippet for the given values?

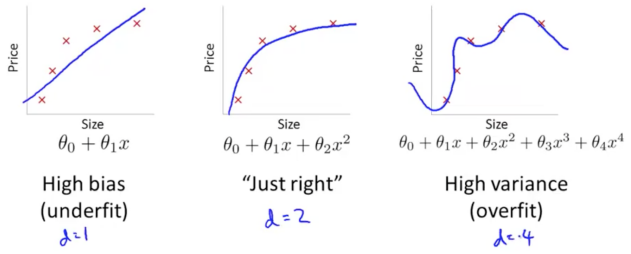
$$M = \begin{pmatrix} 3 & 0 & 2 \\ 2 & 0 & -2 \\ 0 & 1 & 1 \end{pmatrix} v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

```
print M.dot(v)
print v.dot(v)
print v.T.dot(v)
```

1.2.*(6p)* A function named VeriAnaliz on random variable X is defined as follows. $VeriAnaliz_X(x) = P(X <= x \mid x\%2 == 0)$. P indicates the probability of one of the values that the random X variable can take (ie small x). The X variable shows the number of "Head" in tossing a coin three times. Calculate $VeriAnaliz_X(1)$ and $VeriAnaliz_X(2)$?

2. (Model evaluation)

2.1.*(6p)* Explain the concepts of underfitting and overfitting by considering the following figures.



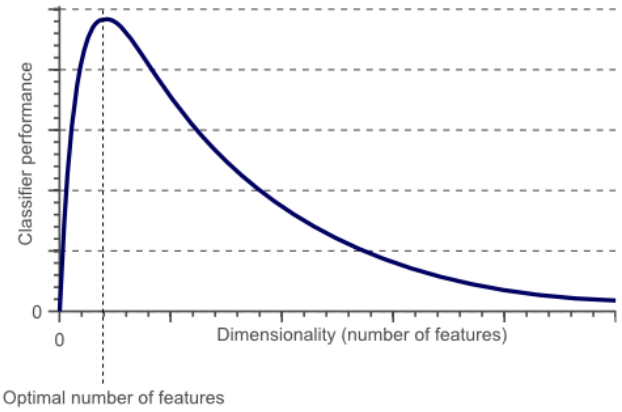
2.2.*(4p)* Why do we evaluate the performance of a model?

2.3.*(10p)* Consider a supervised regression problem with 3000 training samples (samples) each with 100 features, where the features are held in a 3000x100 X matrix and the labels in the 3000x1 by y vector. Suppose you have a model with the k parameter as shown below. The value of k can be a value from 1 to 10. Give the pseudo-code that finds the best value for k. Give the pseudo-code for the accuracy metric of the final model.

- model = train(X, y, k); % Train model on {X, y } with k parameter
- yhat = predict(model, Xhat); % estimate on Xhat using the model

3. (Dimensionality and Feature Selection)

3.1.*(4p)* Interpret the figure?

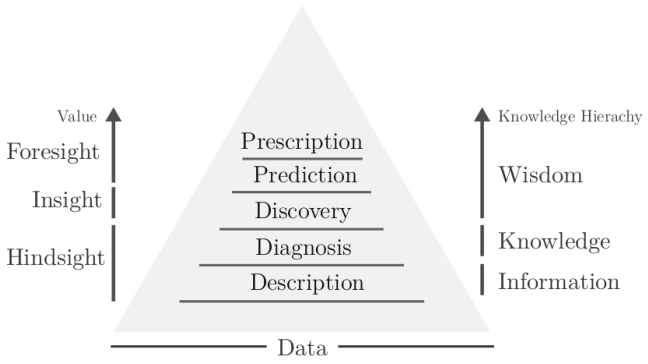


3.2.*(4p)* What is the difference between feature extraction and feature selection, explain it in a sentence and give an example for both.

3.3.*(5p)* How do you decide that a dataset is balanced?

4. (Big data analytics overview)

4.1.*(5p)*Explain the analytical, knowledge, and value hierarchy given below in a maximum of two sentences.



4.2.*(5p)* What is the question to be answered in each data analytics technique, give one question sentence?

Technique name	The question sought an answer in technique

5. (Hadoop Design Patterns)

5.1.*(15p)* Write a Mapreduce pseudocode for finding intersecting rectangular with another one?

5.2.*(10p)* What is a design pattern and give two advantages of its use?

6. (Apache Spark)

(20p) Design a real world Apache Spark project?