MOD500 Decision Analysis with Artificial Intelligence Support

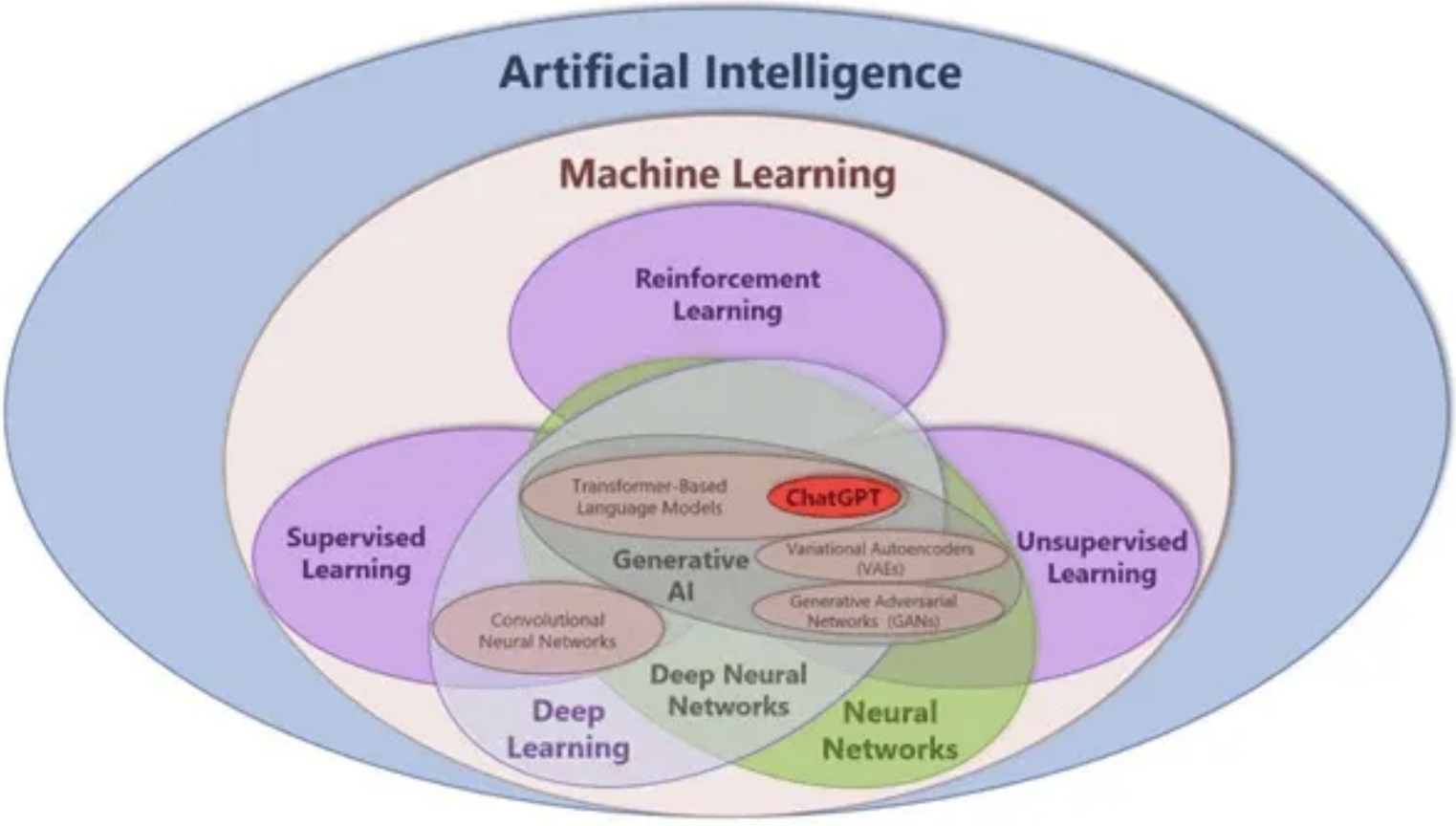
Enrico Riccardi[1]

Department of Energy Resources, University of Stavanger (UiS).[1]

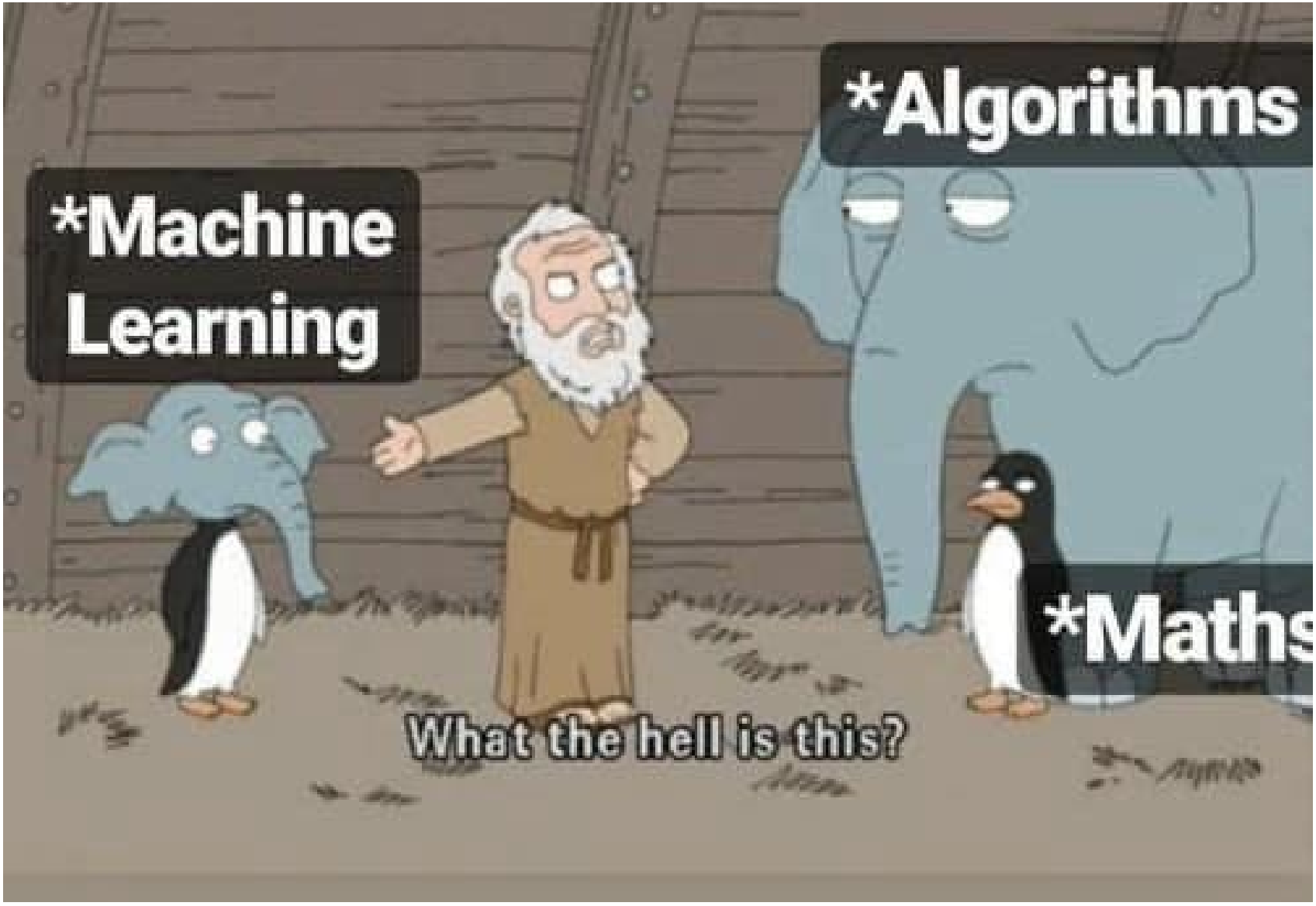
Sep 29, 2024

Statistics, Machine learning or Artificial intelligence?

What is the main difference between the three fields?



How Machine Learning Started?



Statistics

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Statistics

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Statistics

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Statistics

Let's start from the definition

- Statistics (origin "description of a state/country") is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- It is conventional to begin with a statistical population or a statistical model to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal".
- Statistics deals with every aspect of data, including the planning of data collection in terms of the design of surveys and experiments.[Wikipedia]

Machine learning

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]

Machine learning

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]

Machine learning

Definitions:

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. [IBM]
- Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [WIKI]

Machine learning

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Machine learning

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Machine learning

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Machine learning

One technical definition

Machine learning is a set of computer based statistical approaches that aim to minimise the loss function to maximise inference accuracy. [Enrico, 5.2.2024]

The loss function is the actual engine in machine learning.

Loss function

It quantifies the difference between the predicted outputs of a machine learning algorithm and the actual target values.

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

Artificial intelligences

And more definitions:

- Artificial intelligence is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. It is a field of study in computer science that develops and studies intelligent machines. [WIKI]
- It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. [IBM]

Data properties

- All starts from data: what are data-properties?
- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.
- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?
- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.
- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?
- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.
- TRASH in TRASH out

Data properties

- All starts from data: what are data-properties?
- Are there such things as good data and bad data?

Main lesson (Exam question)

- Data DO NOT always have value.

- TRASH in TRASH out

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData properties

Data without metadata are just numbers (i.e. if they are integers, they are still good to play lottery)

Metadata can be pretty much anything. Depending on the application, we can distinguish between:

1. Descriptive : used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
2. Structural : describe how compound objects are put together. It describes the types, versions, relationships, and other characteristics of digital materials.
3. Administrative : to help manage a resource, like resource type, permissions, and when and how it was created.
4. Reference : to indicate the information about the contents and quality of statistical data.
5. Statistical : (or process data), may describe processes that collect, process, or produce statistical data.
6. Legal : creator, copyright, licensing.

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

MetaData for sharing and re-use

More considerations:

- Metadata is more and more important in a digital open world.
- Researchers and automatic algorithms would benefit from importing data directly.
- FAIR research is an important part of Open Science revolution (Findable, Accessible, interoperable, Reusable)
- New applications, business, discoveries can be thus enabled.
- ChatGPT, Bard, Gemini, and all the LLMs are functional only thanks to this!

Super controversial

- Who would be responsible for them then?
- What is the advantage for who releases the data?
- Who gets the money for what?
- Copyright for data and/or for data processing?

Good examples

- Norwegian offshore directorate
- Norway Statistics
- World statistics
- Code repositories
- Data repositories