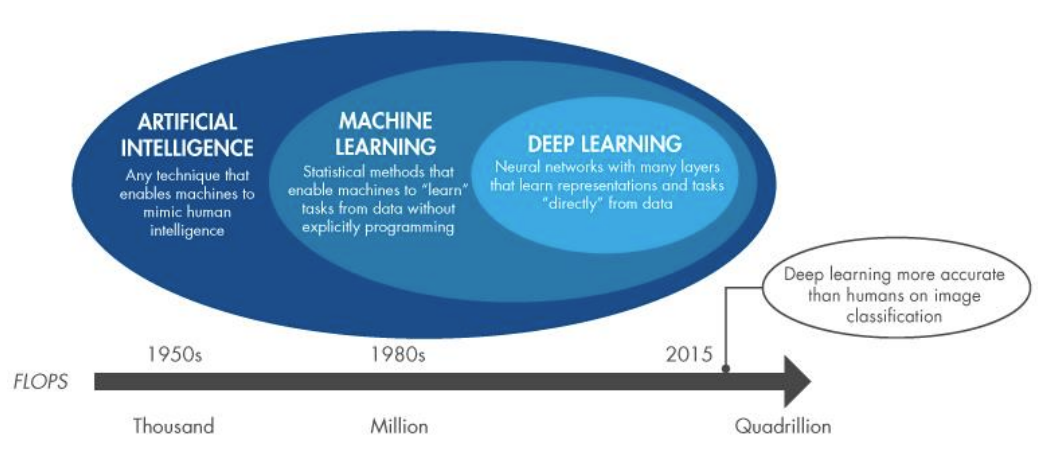
1. **forelesning**



* **AI**
* "The area of computer science that studies how machines can perform tasks that would normally require a sentient agent."
* "The area of computer science that studies how machines can closely imitate human intelligence."
* **ML**
* Machine Learning (ML) is a subset of AI that focuses on the development of algorithms and models that enable computers to learn from data. Instead of being programmed with specific instructions, machines are given access to data and allowed to draw their own conclusions. ML algorithms can be categorized into two main types: supervised learning, where the model is trained on a labeled dataset with predefined outcomes, and unsupervised learning, where the model must find patterns in data without predefined outcomes.
* **DL**
* Deep Learning is a specialized form of machine learning that uses neural networks with multiple layers (deep neural networks) to analyze data. These deep networks are capable of automatically learning to represent data through multiple levels of abstraction. Deep learning has proven particularly effective in areas such as image and speech recognition. Deep neural networks consist of layers of nodes (neurons) that process and transmit information through the network.
* **Understand and explain figure:**

Et bilde som inneholder tekst, skjermbilde, Font, design

Automatisk generert beskrivelse

**Data:**

At the base level, you have raw facts and figures, which constitute data. Data, in its raw form, lacks context and meaning.

**Information:**

Data is processed and organized to form information. Information is meaningful data that provides context. It answers the questions of who, what, where, and when.

**Knowledge:**

When information is analyzed and interpreted, it transforms into knowledge. Knowledge involves understanding the "how" and "why" behind the information. It represents a deeper level of understanding and context.

**Understanding:**

Understanding goes beyond knowledge. It involves the ability to comprehend the implications, relationships, and connections between different sets of knowledge. Understanding is about grasping the significance and implications of information.

**Intelligence:**

Intelligence is the highest level in this hierarchy. It involves the ability to apply knowledge and understanding to new and complex situations. Intelligent systems can learn, adapt, and make decisions based on the information and knowledge they possess.

* **Critical sinking lol**

1. **forelesning**

**Et bilde som inneholder tekst, skjermbilde, Font, diagram

Automatisk generert beskrivelse**

* **ML (steps of building a ML model)**

1. Load the data. (Upload the data to the machine)
2. data preprocessing (missing values, necessary columns, text (Pipeline), image, sound, numbers)
3. split the data into training and testing sets (supervised learning)
4. Standardization (feature scaling, make the data consistent, 5. Avenue, Avn, Av, Number, nr. N.)
5. Using algorithms to build it
6. Evaluation
7. Improvements

* **Feature selection (Remember one of those techniques)**

For a model to produce accurate results, you need to make sure it’s using the right data. Feature selection is how you ensure your model is focused on the data with the most predictive power and is not distracted by data that won’t impact decision making. Precise feature selection will result in a faster, more efficient, more interpretable model.

* If you have a lot of domain knowledge, use machine learning and manually select the important features of your data.
* If you have limited domain knowledge, you can try to use automatic selections algorithms such as decision trees, random forests, K-nearest Neighbors(KNN) and so on.
* If your data has lots of features, use component analysis to reduce the dimensions of the data.

**Techniques:**

* **Feature Importance** \*Use ExtraTreesClassifier & matplotlib

Provide a score for each feature in a dataset, can be used for selecting important features.

* **Univariate Selection** \*Use SelectKBest & chi2

Provide a score for each feature in a dataset, can be used to determine which features have the strongest correlation to the output variable.

* **Correlation Heatmaps** \*Import seaborn & matplotlib

Provide a matrix to show the relationship between the different values of the features. A heatmap makes it easy to identify which features are more correlated to the target variable.

* **Scaling**

****

A technique for standardizing the range of features in a dataset. The size of a house can be 100 M3 while number of bedrooms might be 3. If we leave these values alone, the

features with a higher scale might be given a higher weighting if left alone. This can be fixed using normalization or standardization.

**Normalization** (minMax) scales all values for a feature within a fixed range between 0 and 1.

Et bilde som inneholder tekst, Font, skjermbilde

Automatisk generert beskrivelse

**Standardization** (Z-score) scales data to have a mean (μ) of 0 and a standard deviation (σ) of 1 (+1,-1).

Et bilde som inneholder tekst, skjermbilde, Font

Automatisk generert beskrivelse

1. **forelesning**

* ML =  
  - Supervised learning (classification, regression) (Classification = KNN, Naïve bayes, Decision tree, Random forrest, RNN/ CNN) (Regression = Logistic, Linear)   
  - Unsupervised learning (Clustering = K-means (customer segmentation))  
  - reinforcement
* Learn one algorithm for each + one application and one limitation.
* Confusion matrix (explain and understand) evaluation approach (True positive is most important prediction) Recall???

**Supervised:**

* Labeled data example one dataset of dogs and one of cats. The machine knows that these two datasets contain different types of animals.

**Classification:**

* **K-Nearest Neighbors (KNN):** A simple algorithm that classifies data points based on the majority class of their k-nearest neighbors.
* **Naive Bayes:** A probabilistic algorithm based on Bayes' theorem that makes assumptions about the independence of features. (Prices of houses)
* **Decision Tree:** A tree-like model where each internal node represents a decision based on the value of a feature, and each leaf node represents the predicted output.
* **Random Forest:** An ensemble learning method that builds multiple decision trees and merges their predictions to improve accuracy and control overfitting.
* **Recurrent Neural Networks (RNN):** A type of neural network designed for sequence data, allowing information persistence through time. (Autocorrect, speech recognition) RNN memorizes previous inputs and can predict outcomes, such as in autocorrect.
* **Convolutional Neural Networks (CNN):** Especially effective for image processing, CNNs use convolutional layers to automatically and adaptively learn spatial hierarchies of features. (image classification (cars and dogs are not the same category), object detection, Image segmentation). CNN can be applied in health care example identify abnormalities in x-rays.

**Regression:**

* **Linear Regression:** A simple linear model that establishes a relationship between the dependent variable and one or more independent variables.
* **Logistic Regression:** Despite its name, logistic regression is used for binary classification problems, predicting the probability that an instance belongs to a particular class.

**Unsupervised Learning:**

* Unlabeled, example animals where the machine must find the difference on its own.

**Clustering:**

* **K-Means:** An iterative algorithm that partitions data into k clusters based on similarity, where each cluster is represented by its centroid.
* **Customer Segmentation:** An application of clustering to group customers based on common characteristics, helping businesses tailor marketing strategies.

**Reinforcement Learning:**

* **Reinforcement Learning (RL):** An area of machine learning where agents learn to make decisions by interacting with an environment. The agent receives feedback in the form of rewards or penalties based on the actions it takes.
* Reinforcement learning is often used in scenarios where an agent learns to navigate and make decisions in complex, dynamic environments, such as in game playing, robotics, and autonomous systems. The agent learns by trial and error, adapting its strategy based on the feedback received from the environment.

**Confusion matrix:**

* A confusion matrix is a table used in classification to evaluate the performance of a model. It compares the predicted classes of a model against the actual classes in the dataset. The matrix is organized as follows:

Et bilde som inneholder tekst, skjermbilde, Font

Automatisk generert beskrivelse

* **True Positive (TP):** Instances where the model correctly predicted the positive class.
* **True Negative (TN):** Instances where the model correctly predicted the negative class.
* **False Positive (FP):** Instances where the model incorrectly predicted the positive class when it was actually negative (Type I error).
* **False Negative (FN):** Instances where the model incorrectly predicted the negative class when it was actually positive (Type II error).

**Evaluation Approach:**

* **Recall (Sensitivity, True Positive Rate):**
* Recall is a metric that assesses the ability of a classification model to capture and correctly identify all instances of the positive class. It is calculated as the ratio of true positives to the sum of true positives and false negatives:
* **Importance of True Positives:**
* In many applications, especially where the cost of false negatives is high (e.g., in medical diagnoses), true positives are crucial. For example, in a medical test for a severe disease, missing a positive case (false negative) can have serious consequences, making recall a vital metric.
* **Trade-off with Precision:**
* Recall is often in tension with precision. Increasing recall might lead to more false positives, reducing precision, and vice versa. Achieving the right balance depends on the specific goals and constraints of the problem at hand.
* **F1 Score:**
* The F1 score is the harmonic mean of precision and recall and provides a balance between the two. It is especially useful when the class distribution is imbalanced.

Understanding and optimizing recall is crucial when dealing with scenarios where the cost of missing positive instances is high, and correctly identifying all positive cases is of utmost importance. It helps in ensuring that the model has a high sensitivity to the target class.

1. **forelesning**

* Know the difference between supervised and unsupervised learning.
* Supervised = labeled data, x selected features and y targeted features. regression
* Unsupervised = unlabeled data, clustering

**Supervised Learning:**

**Definition:**

* In supervised learning, the algorithm is trained on a labeled dataset, where each input is associated with the corresponding correct output. The goal is to learn a mapping from input to output, and the algorithm aims to make predictions or decisions based on this learned mapping.
* **Key Characteristics:**
* Labeled Data: The training dataset includes both input features and their corresponding correct output labels.
* **Goal:**
* To learn a mapping function that can generalize well to new, unseen data.
* **Training Process:**
* The algorithm is trained by adjusting its parameters based on the error (difference between predicted and actual output) to minimize this error.
* **Common Applications:**
* Classification and Regression problems, where the algorithm predicts a label or a numeric value based on input features.
* Examples include image classification, spam detection, and predicting house prices.

**Unsupervised Learning:**

**Definition:**

* In unsupervised learning, the algorithm is given unlabeled data and is left to discover patterns, structures, or relationships within the data on its own. The objective is often to find inherent structures or groupings without explicit guidance on the correct output.
* **Key Characteristics:**
* Unlabeled Data: The training dataset does not include output labels; the algorithm explores the data's inherent structure without predefined categories.
* **Goal:**
* To discover patterns, clusters, or relationships within the data.
* **Training Process:**
* The algorithm explores the data and identifies similarities, differences, or structures without explicit guidance on the correct output.
* **Common Applications:**
* Clustering, Dimensionality Reduction, and Association problems.
* Examples include customer segmentation, anomaly detection, and topic modeling.

**Summary of Differences:**

* **Guidance:**
* Supervised Learning: Guided by labeled examples with input-output pairs.
* Unsupervised Learning: Explores unlabeled data to find patterns without explicit guidance.
* **Goal:**
* Supervised Learning: Predict output labels or values for new, unseen data.
* Unsupervised Learning: Discover patterns, structures, or relationships within the data.
* **Examples:**
* Supervised Learning: Classification, Regression.
* Unsupervised Learning: Clustering, Dimensionality Reduction.

In practice, the choice between supervised and unsupervised learning depends on the nature of the problem and the type of information available in the dataset.

1. **forelesning**

* chatbots = examples, how can it be applied and used; capabilities of chatbots

**Examples of Chatbots:**

* **Customer Support:**
* Many companies use chatbots to provide instant support to customers, answering common queries, troubleshooting issues, and guiding users through processes.
* **E-commerce:**
* Chatbots in e-commerce platforms assist users in finding products, placing orders, and answering questions about shipping and returns.
* **Virtual Assistants:**
* Personal virtual assistants like Siri, Google Assistant, and Alexa leverage chatbot technology to understand and respond to user voice commands, perform tasks, and provide information.
* **Booking and Reservations:**
* Chatbots facilitate booking services, such as hotel reservations, flight bookings, and restaurant reservations, streamlining the process for users.
* **Healthcare Support:**
* Chatbots in healthcare assist users in scheduling appointments, providing information about medications, and offering basic medical advice.

**Applications and Use Cases:**

* **Automated Customer Interaction:**
* Chatbots handle routine customer interactions, saving time for both customers and businesses. They can answer frequently asked questions, guide users through processes, and provide product information.
* **Lead Generation:**
* Chatbots on websites can engage with visitors, collect information, and qualify leads. They can assist in the initial stages of the sales funnel.
* **Employee Assistance:**
* Internal chatbots can help employees by answering HR-related queries, providing information about company policies, or guiding them through onboarding processes.
* **Language Translation:**
* Chatbots can be equipped with language translation capabilities, facilitating communication between users who speak different languages.
* **Education and Training:**
* Chatbots in educational settings can assist students with homework, provide study resources, and offer guidance on course selection.

**Capabilities of Chatbots:**

* **Natural Language Processing (NLP):**
* Chatbots use NLP to understand and interpret human language, allowing them to respond in a conversational manner and comprehend user intent.
* **Context Awareness:**
* Advanced chatbots can maintain context during a conversation, remembering previous interactions and providing more personalized and relevant responses.
* **Multi-Channel Integration:**
* Chatbots can be integrated across various communication channels, including websites, messaging apps, and social media platforms, ensuring a consistent user experience.
* **Task Automation:**
* Chatbots automate repetitive tasks, reducing the need for human intervention in routine processes.
* **Learning and Adaptation:**
* Some chatbots employ machine learning algorithms to learn from user interactions over time, continuously improving their ability to understand and respond effectively.
* **Security Measures:**
* Chatbots are designed with security features to handle sensitive information securely, ensuring user data privacy and compliance with regulations.
* **Integration with External Systems:**
* Chatbots can connect with databases, APIs, and other systems to retrieve and provide real-time information, enhancing their functionality.

Chatbots have evolved to become versatile tools with applications across various industries, providing efficient and personalized interactions for users. Their capabilities continue to advance with advancements in natural language processing, machine learning, and artificial intelligence.

1. **forelesning**

* NLP (Natural Language Processing)
* Pipeline = understand it
* Understand how to build an NLP model (page 30)

NLP is about making computers able to understand human language.

Et bilde som inneholder tekst, skjermbilde, Font, design

Automatisk generert beskrivelse

**Sentiment Analysis:**

* Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves determine the sentiment or emotional tone expressed in a piece of text, such as a movie review, tweet, or comment.
* The primary goal of sentiment analysis is to understand and classify the sentiment as positive, negative, or neutral.

**Build an NLP model:**

* **Load the data:**
* Give the machine access to a dataset.
* **Text preprocessing (Remember the steps in ML):**
* Handle missing values, remove unnecessary columns AND Pipeline process.
* **Split the data into training and testing sets:**
* 80% train and 20% test.
* **Standardization (Transform text into numerical data):**
* Feature extraction in Pipeline process.
* Popular word vectorizers include Word2Vec, GloVe, FastText.
* See pictures under for visualization.
* **Train an algorithm:**
* For example, Naive Bayes classifier.
* **Evaluate the model.**
* **Save the trained model.**

Et bilde som inneholder tekst, skjermbilde, Font, nummer

Automatisk generert beskrivelse

Et bilde som inneholder tekst, skjermbilde, Font

Automatisk generert beskrivelse

1. **forelesning**

* examples for several applications of computer visions (CV)
* explain at least one example and how it works (ML model building steps)
* understand pixels and RGB channels
* remember one algorithm (SIFT, SURF, ORB)

**Object detection:**

* Identifying and locating objects with images or video frames.
* Commonly used in cars for recognizing pedestrians, traffic signs and other cars.

**Object tracking:**

* VR headset position tracking, hand and body 3D 360 cameras.
* Monitoring and following movements of objects over time.
* Used for VR, surveillance systems and tracking general movement.

**Object verification:**

* Confirming if the object is what it’s expected to be.
* Can be used for facial recognition or finger scanning.

**Object identification:**

* Categorizing the detected object into different categories.
* Can be used for image classification where it can sort pictures into categories such as mushrooms, dog breeds, plants and so on.

These can be used for picture analysis, face detection, 3D modelling and rebuilding, robotics, optional character recognition (OCR) for pictures with text in them.

You can use convolution neural networks (CNN) which automatically learns and extract hierarchical structures from images.

**Pixels and RGB channels:**

* Pixel features refer to characteristics or attributes associated with individual pixels in an image.
* Pixel features are typically represented as numerical values.
* Pixel features are the result of image transformation, this process can be achieved by algorithms.
* Each pixel in an image is associated with a set of numerical attributes or characteristics that describe various properties of that pixel, such as its color intensity, position, and so on. For example, in grayscale images, the pixel's intensity is represented by a single numerical value ranging from 0 (black) to 255 (white).
* RGB are red, green and blue colors. Used in colored pictures and the numeric value of each pixel is calculated by adding the intensity of each color in that pixel and divided by 3.

Et bilde som inneholder skjermbilde, sort og hvit, kvadrat, mønster

Automatisk generert beskrivelse

Et bilde som inneholder tekst, diagram, line, skjermbilde

Automatisk generert beskrivelse

**Algorithms:**

* **SIFT (Scale-invariant feature transform):**
* SIFT is an algorithm that extracts key points and their descriptors from an image. It is known for its scale and rotation invariance, making it robust to changes in an object's size and orientation.
* Key Points: SIFT detects distinctive local features, such as corners and blobs, and computes descriptors that capture information about the surrounding region's texture and gradient.
* **SURF (Speeded-Up robust features):**
* Detecting and describing key points in images, faster than SIFT while having high performance.
* **ORB (Orientated FAST and rotated BRIEF):**
* ORB is a fast and efficient algorithm for feature detection and description. It is designed to be computationally lightweight and suitable for real-time applications.
* Key Points: ORB combines the FAST (Features from Accelerated Segment Test) corner detector with the BRIEF (Binary Robust Independent Elementary Features) descriptor to extract and match key points.

1. **forelesning**

* Neural Networks, Explain picture 9 in ppt. Input layers, neurons and channels, output layers = predicted results.
* Know components of Neural Network (NN) page 12.

Et bilde som inneholder skjermbilde, tekst

Automatisk generert beskrivelse

* **Neurons:**
* Nodes that process and transmit information.
* **Layers:**
* Organized into input, hidden and output layers.
* **Weight:**
* Connections between layers of neurons.
* Adjusted during training to minimize errors.
* **Learning:**
* The process of adjusting weight to improve predictions.
* **backpropagations:**
* Algorithm for updating weights during training.
* **Activation functions:**
* Determining if neurons are activated or not.

1. **Forelesning**

* Deep learning (Only one question)
* CNN vs RNN
* CNN = Images
* RNN = sequential data (text, speech, voice)
* Unstructured data
* Understand the case study (page34), what is cat and what is dog.
* Les hele PPT!!!

**Deep learning:**

* Is a subfield of machine learning that focuses on the development of artificial neural networks, inspired by the structure and function of the human brain, to solve complex tasks.
* It is a type of machine learning where algorithms attempt to learn and make decisions directly from data, without relying on explicit programming.
* Deep learning models are typically built using artificial neural networks, which consist of interconnected layers of artificial neurons or "nodes."

**CNN (convolutional neural network):**

* Are primarily used for image data. The algorithm is designed to do task such as image classification, object detection and image segmentation.
* CNN use convolutional layers to learn features from images which makes them highly used for computer vision tasks. It is commonly used in facial recognition, picture classification, helping doctors with detecting abnormalities in x-rays, diagnose plants and much more.

**RNN (recurrent neural network):**

* Are designed for preprocessing sequential data. These are texts, speech and voice and more. Its commonly used to voice assistants, suggesting words when typing, speech recognition, predicting patient outcomes and so on. They are basing their decisions on the data coming in. The idea is that the algorithm memorizes the previous inputs and use them for the future. This makes the algorithm perfect for tasks such as text to speech, autocorrect etc. as it previously has encountered the same speech pattern.

They both take in unstructured data as the data can be images, speech or random text. As an example, they are used in the word suggestion on mobile devices that often can be annoying if it’s always wrong or makes no sense based on your dialect. But they are great tools for those who find them useful.

Et bilde som inneholder tekst, skjermbilde, Font

Automatisk generert beskrivelse

Et bilde som inneholder tekst, skjermbilde, Font, dokument

Automatisk generert beskrivelse

Et bilde som inneholder tekst, skjermbilde, Font, nummer

Automatisk generert beskrivelse

1. **Forelesning**

* Ethical issues
* Have a look at some current ongoing debate
* We will be given several cases, and I should give arguments for pros and cons

**Current ongoing debates:**

* Deepfakes, bias and fairness, job replacements, ethical use, customer support, school and education, lethal autonomous weapons.