



Azure AI

AI Workloads

- **Anomaly Detection**
 - It is the process of finding outliers within a dataset called an anomaly
- **Machine Learning**
 - It's process of machine that gets better at a task without explicit programming.
- **Computer Vision**
 - when we use ML or NN to gain high-level understanding from digital images or video
- **Natural Language Processing (NLP)**
 - ML that can understand the context of a corpus (a body of related text)
- **Conversational AI**
 - is technology that can participate in conversations with humans.
- **Knowledge mining**
 - is a discipline in AI that uses a combination of intelligent services to quickly learn from vast amounts of information. You can use Azure Cognitive Search's knowledge mining results and populate your knowledge base of your chatbot.

Principles of Responsible AI

- a. Fairness
- b. Reliability
- c. Privacy & Security
- d. Transparency
- e. Accountability
- f. Inclusiveness

1. Fairness: AI systems should treat all people fairly

- Fairness is a core ethical principle that all humans aim to understand and apply. This principle is even more important when AI systems are being developed. Key checks and balances need to make sure that the system's decisions don't discriminate or run a gender, race, sexual orientation, or religion bias toward a group or individual.

2. Reliability and Safety: AI systems should perform reliably and safely

- To build trust, it's critical that AI systems operate reliably, safely, and consistently under normal circumstances and in unexpected conditions. These systems should be able to



operate as they were originally designed, respond safely to unanticipated conditions, and resist harmful manipulation.

- To build trust, it's critical that AI systems operate reliably, safely, and consistently under normal circumstances and in unexpected conditions. These systems should be able to operate as they were originally designed, respond safely to unanticipated conditions, and resist harmful manipulation.
- AI systems need to be reliable and safe in order to be trusted. It is important for a system to perform as it was originally designed and for it to respond safely to new situations. Its inherent resilience should resist intended or unintended manipulation. Rigorous testing and validation should be established for operating conditions to ensure that the system responds safely to edge cases, and A/B testing and champion/challenger methods should be integrated into the evaluation process. An AI system's performance can degrade over time, so a robust monitoring and model tracking process needs to be established to reactively and proactively measure the model's performance and retrain it, as necessary, to modernize it.

3. Privacy and Security: AI systems should be secure and respect privacy

- A data holder is obligated to protect the data in an AI system, and privacy and security are an integral part of this system. Personal needs to be secured, and it should be accessed in a way that doesn't compromise an individual's privacy.
- As AI becomes more prevalent, protecting privacy and securing important personal and business information is becoming more critical and complex. With AI, privacy and data security issues require especially close attention because access to data is essential for AI systems to make accurate and informed predictions and decisions about people. AI systems must comply with privacy laws that require transparency about the collection, use, and storage of data and mandate that consumers have appropriate controls to choose how their data is used

4. Inclusiveness: AI systems should empower everyone and engage people

- At Microsoft, we firmly believe everyone should benefit from intelligent technology, meaning it must incorporate and address a broad range of human needs and experiences. For the 1 billion people with disabilities around the world.
- Inclusiveness mandates that AI should consider all human races and experiences, and inclusive design practices can help developers to understand and address potential barriers that could unintentionally exclude people. Where possible, speech-to-text, text-to-speech, and visual recognition technology should be used to empower people with hearing, visual, and other impairments.



5. Transparency: AI systems should be understandable

- Achieving transparency helps the team to understand the data and algorithms used to train the model, what transformation logic was applied to the data, the final model generated, and its associated assets. This information offers insights about how the model was created, which allows it to be reproduced in a transparent

6. Accountability: People should be accountable for AI systems

- The people who design and deploy AI systems must be accountable for how their systems operate. Organizations should draw upon industry standards to develop accountability norms. These norms can ensure that AI systems are not the final authority on any decision that impacts people's lives and that humans maintain meaningful control over otherwise highly autonomous AI systems.

DATA SCIENCE
by
SIVA RAMA KRISHNA



Cognitive Services

Cognitive Services is an umbrella AI service that enables customers to access multiple AI services with an API key and an API Endpoint

- **Anomaly Detector** - Identify potential problems early on.
- **Content Moderator** - Detect potentially offensive or unwanted content.
- **Personaliser** - Create rich, personalized experiences for every user.

NLP Services

- **Language Understanding** - Build natural language understanding into apps, bots and IoT devices.
- **QnA Maker** - Create a conversational question and answer layer over your data.
- **Text Analytics** - Detect sentiment, key phrases and named entities.
- **Translator** - Detect and translate more than 90 supported languages.
- **Speech to Text** - Transcribe audible speech into readable, searchable text.
- **Text to Speech** - Convert text to lifelike speech for more natural interfaces.
- **Speech Translation** - Integrate real-time speech translation into your apps.
- **Speaker Recognition** - Identify and verify the people speaking based on audio.

CV Services

- **Computer Vision** - Analyze content in images and video.
- **Custom Vision** - Customize image recognition to fit your business needs.
- **Face** - Detect and identify people and emotions in images.
- **Form Recognizer** - Accelerate your business processes by automating information extraction. Form Recognizer applies advanced machine learning to accurately extract text, key/ value pairs, and tables from documents. With just a few samples, Form Recognizer tailors its understanding to your documents, both on-premises and in the cloud. Turn forms into usable data at a fraction of the time and cost, so you can focus more time acting on the information rather than compiling it.



Data

Structured data (tabular data) – Multiple Columns & Multiple Rows

- Columns → Variables or Features
- Rows → Records

Input Variables → Independent Variables [1,inf)

Output Variable → Dependent Variable or Response (1)

If you have output Variable (this dataset is called as **labeled data**) & you have to predict the output variable based on input variables, then we to apply **Supervised Machine Learning**.

- ➔ Regression (output → Numeric)
- ➔ Classification (output → classes/categories)

If output Variable is not available (this dataset is called as unlabeled data) & you have to predict based on input variables, then we to apply **Unsupervised Machine Learning**

- ➔ Clustering
- ➔ Association Rules
- ➔ Recommendation Engines

Unstructured Data – Images, Videos, Text, Audio

- ➔ Images & Videos : Computer Vision (CV)
- ➔ Text & Audio : Natural Language Processing (NLP)



Machine Learning Project (pipeline-Step by step execution)

1. Business Problem Understanding
2. Data Understanding & Dataset Understanding
3. Data Preprocessing
 - Feature Engineering
 - Data Cleaning
 - Data Wrangling
 - Dimension Reduction (PCA/Statistics)
 - Train test split
 - i. Train data → Machine to learn
 - ii. Test Data or Validation Data → to evaluate
4. Model → Applying ML algorithm on train data
5. Evaluation / Validation → Check ML Model performance on Test data
 - i. Evaluation Metrics for Regression
 1. Mean Absolute Error
 2. Mean Square Error
 3. Root Mean Square Error
 4. Coefficient of Determination (R^2)
 - ii. Evaluation Metrics for Classification
 1. Accuracy
 2. Confusion Matrix
 3. Precision
 4. Recall
 5. F1-Score
 6. True Positive Rate (TPR) → Sensitivity → Recall (+ve)
 7. True Negative Rate (TNR) → Specificity → Recall (-ve)
 8. False Positive Rate (FPR) = 1-TNR
 9. False Negative Rate (FNR) = 1-TPR
 - iii. Clustering
 1. Measures (Within Cluster Sum of Squares Distances)



ML Pipeline

- Pre Processing : preparing data and feature engineering before passing data to an ML model for training or inference
- Data cleaning - correcting errors within the dataset that could negatively impact the results
- Data reduction - reducing the amount of data, or applying dimensionality reductions to reduce the dimensions of input vectors
- Feature engineering - transforming data in numerical (vectors) to be ingested by an ML model
- Sampling - balancing a dataset to be uniform across labels by adding or removing records
- Post Processing - translating the output of an ML model back into a human readable format
- Training the process of training the model
- Serving the the process of deploying a model to an endpoint to be used for inference
- Inference Invoking an ML model by sending a request and expecting back a prediction
- Real-time endpoint - optimized for small or single item payloads, returns results quickly (a dedicated running server)
- Batch-transform endpoint - optimized for larger batch predictions (server runs only for the duration of the batch)
- Forecasting: Makes a future prediction with relevant data, analysis of trends.
- Predicting: Makes a future prediction without relevant data, uses statistics to predict future outcomes, more of "guessing", Uses decision theory
- Performance/Evaluation Metrics are used to evaluate different Machine Learning Algorithms
- Classification is a process of finding a function to divide a labeled dataset into classes/categories
- Classification Metrics (Accuracy, F1 Score, Precision, Recall)
- Confusion matrix is table to visualize the model predictions (predicted) vs ground truth labels (actual)
- Regression is a process of finding a function to correlate a labeled dataset into continuous variable/number.
- Regression Metrics (MSE, RMSE MAE)
- Clustering a process grouping unlabeled data based on similarities and differences.



Natural Language Processing

Sentiment analysis find out what people think of your brand or topic

Ex: Labels include negative, positive, mixed or neutral

granular data with a Subject and Opinion tied to a Sentiment

key phrase extraction : quickly identify the main concepts in text.

language detection : detect the language an input text is written in

Named entity recognition (NER) : detects words and phrases mentioned in unstructured text that can be associated with one or more semantic types.

entities-what parts of the intent is used to determine the answer

Language Understanding (LUIS) : is a no-code ML service to build natural language into apps, bots, and IoT devices

Natural Language Understanding (NLU) the ability to transform a linguistic statement to a representation that enables you to understand your users naturally

QnA Maker generate a bot from a URL, PDF or DOX

Azure Bot Service - allows you to host bots