Welcome to:

Applications of Machine Learning in Government Administration

Unit objectives

After completing this unit, you should be able to:

- Understand the type of government problems appropriate for AI applications
- Learn some citizen service use cases
- Understand the importance of chat boots for communication between citizen and government
- Gain knowledge on applications of machine learning techniques to understand the sentiment analysis of citizen
- Learn about machine learning for classification of citizen complaints

Introduction

- For a range of applications in the investment management sector, Artificial Intelligence (AI) and machine learning are increasingly being implemented.

- Some of AI and machine learning's current and potential use cases include:
  - AI and machine learning approaches are used by financial institutions and suppliers to determine credit performance, cost and business insurance contracts and to automate customer interaction.
  - Institutions leverage scarce resources with AI and machine learning methods, as well as back-testing models, and evaluate the market impact of large-scale trading.
  - Hedge funds, broker dealers and other companies use AI and machine learning to identify triggers for higher (and uncorrelated) returns and boost trading efficiency.
  - Such tools may be used by both public and private sector entities for regulatory compliance, reporting, data quality evaluation and fraud detection.

Risk and compliance

- Cognitive systems can improve the visibility of regulatory and internal compliance controls throughout the company.

- Through knowing the current regulatory laws and matching bank operation with them, cognitive systems can help bank risk officers identify when a bank is unable to fulfil their obligations.

- The bank has implemented cognitive technology to reduce their exposure to risk and strengthen their due diligence processes.

Type of government problems appropriate for AI applications

Shortage of Potential Human resource	• **Poor Administrative support due to lack of potential human resource.** • **Enquire response time are too long due to insufficient support.**
Large Data sets	• Difficult to work efficiently with too large public datasets. • Lack of technologies and skill sets to make advantage of available data. • Facing Challenges in maintain structured data due to lack of technical interventions. • Statistical and predictive data should be frequently described
Shortage of exports and expertize	• It is possible to tackle simple issues, saving up time for experts. • It is important to study specialized problems to help analysis specialists
Predicting the scenarios	• Condition on the basis of empirical information is reliable. • Precision helps with moment-sensitive reactions
Repetitive tasks	• Most of the government administration tasks and procedures which are repetitive can be replaced by ML models.

AI for citizen services use cases

- Government offices around the world are researching AI systems.
- Public enquiries and information related to the prevalent public programs usage cases.
- AI can dramatically improve citizen’s exposure to actual period responses, and can even be utilized to produce and complete reports, particularly for repetitive activities.
- AI will also have greater effect if it genuinely reduces organizational pressures and increases user service rather than substituting staff.
- AI systems, when implemented strategically, could deliver community services more effectively while potentially reducing price and rising resident’s engagement and commitment.

AI use cases for citizen inquires and information

- Answering questions.
- Routing requests.
- Translation.
- Drafting documents.

Answering questions

- At the North Carolina state department, catboats auditory or text-based automated communicative services that are often AI-based safe the operator's assistance centre line.

- The purpose of these catboats is to act as virtual members.

- In the upcoming years, AI can also be utilized to interpret the emotions of applications and forums in order to better understand citizen’s inquiries and suggestions.

Routing requests

- Effort to utilize software to recognize and channel the requests of citizens to the appropriate agency.

- To supply consumers with fundamental queries regarding their coverage and expenses and a path to their closest workplace.

Translation

- Many programs now make it possible to interpret government information more effectively.

- Unable incorporates crowdsourcing and machine learning in 14 languages to translate business transactions.

Drafting documents

- Export and production is crewing a program to assist legislators in reacting to citizen's queries by producing AI responses.

- Documentation can be achieved using Natural Language Generation (NLG) AI, often utilized in hundreds of media outlets.

Chat bots for communication between citizen and government

- The initial version of chatbots has minimal capacity to react to simple law-based inquiries.

- Cost savings, reduction of the workload of public servants, improved efficiency, generation of new job prospects, solving asset management issues, quality of public activity and enhancement of the experience of people were widely reported.

- The implementation of public-sector chatbots introduces a new research field that raises a set of issues concerning the application of AI technologies and its approach to the distribution of public services.

Media richness theory

- It is most commonly known and utilised concepts of interaction channels, offering a valuable correlation among the features of a interaction role to be carried out with the features of the medium were using for it.

- The key discovery of this study was that portion of legal resident-government communication is plain, primarily about knowledge desiring or performing low-complexity, unclear and uncertain accounts.

Chatbots in the public sector

- A growing range of AI implementations in society are associated with chatbots or communicative assistants.

- As the national industry seeks ways to enhance citizen's and government facilities and processes, more advanced AI implementations have emerged, primarily aimed at the automated delivery of government information.

- Recorded government AI programs primarily focus on supporting consumer (citizen) provider through trolls that provide answers to simple citizen inquiries and appropriate data.

- A study on AI and potential democracy Centre for people Intervention, discusses four non-entire AI skills that can improve the cycle of social service process:
  - Predictive Analysis (AP).
  - Prediction.
  - Virtual Perception (CV).
  - Application of common communication (NLP).

Enhancement of bots intelligence

- E-government systems are correlated with large, ever-improving quantity of various forms of information from various and dispersed outlets.
- Recent solutions to these criteria focus on individual and computer cognition interaction.
- Creative document processing and artificial intelligence strategies enable this to evaluate the meaning of unorganized information such as statements and records in the human word.
- It will involve a change in emphasis from pure data compilation and depiction to its practical analysis, collection, optimizing and in modern e-state methods.
- The latter request for a creative method based on:
  - Smart textual analysis, structuring and aggregating weighty and complicated information.
  - Insightful study and manipulation of information trends and interconnections.
  - Catching shareholder's tacit awareness of information interpretation and problem solving through a social network method.
  - Leveraging the characteristics overall.

Case study

- Interaction between government and citizen:
  - The system allows heterogeneous systems to be seamlessly integrated and guarantees their interoperability from a point of view of technological, theoretical and user experience.
  - The approach anticipated is universal and has no back-end technology constraints.
  - The solution makes it easy to access integrated resources and streamlines a variety of related needs.

Data management services

- This allows:
  - Focused exploration.
  - Collection.
  - Indexing.
  - Exchange.
  - Storage in a variety of data sources and forms of solvable huge-scale content.

- Information processing systems are designed semantic to efficiently resolve the relevant schema and epistemology issues.

Knowledge processing services

- It benefits from the most popular large-scale information storage technology to deliver functionality such as:
  - Low quality data mining.
  - Focused data archiving.
  - Filtering.
  - Grouping.
  - Aggregation.
  - Analysis.

- The suggested content strategies will better operate the above-mentioned information in order to recognize essential pieces of knowledge and hidden patterns and link them substantively with the needs and choices of real people.

- Data gathering is based on several well-established methodologies and methods in the suggested solution, such as:
  - Cognitive systems.
  - K-means.
  - Trees judgment.

Application services

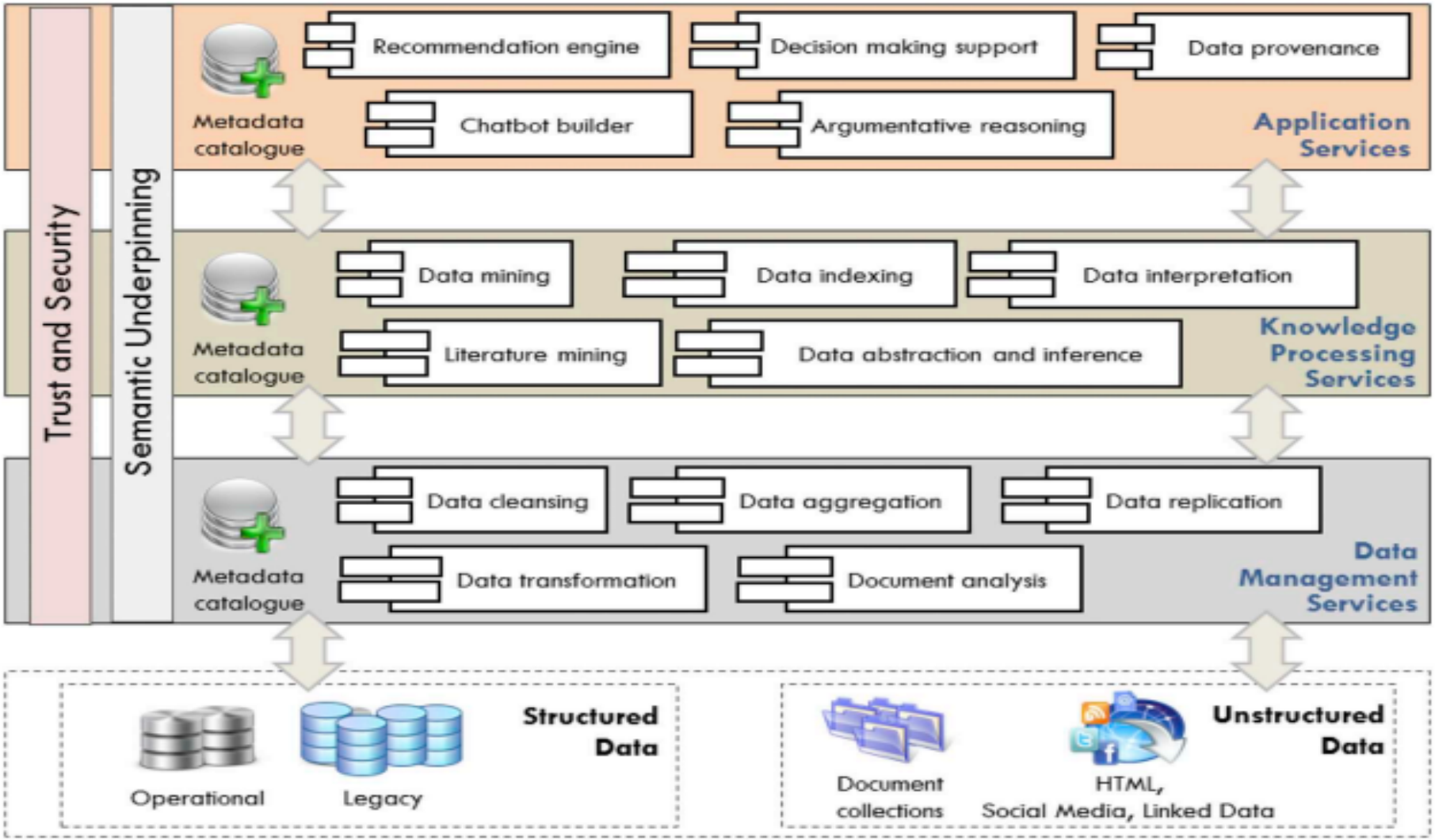

Figure: Application services

Source: http://www.icsd.aegean.gr/publication_files/Journal/980795393.pdf

An application scenario

- Computerized data about local events and news, provincial points and services, decisions and deadlines for public transit, housing and restaurants, other tourist data, etc. can be provided by trolls for the former provider classification.

- If the suggested approach is used to offer feedback on "retirement pension requests“, this scenario applies to the bot-to-citizen interaction workflow.

- The individual initially uses the chatbot's state-operated communicative framework via an authenticated authentication process.

- The chatbot then asks a series of queries that the choice-making support service will answer in attempt to evaluate the relevance of the circumstances for their pension implementation.

- Such an option is focused on a set of guidelines taken from the applicable regulatory records.

- The network's data management systems use the appropriate social media APIs to access that information using password hashes, thanks to the client's consent.

Classifications of citizen complaints using ML

- To improve the handling of problems in society, social medias uses many sources of information.

- Twitter (social media) has become a famous cultural network based on micro blogging that enables users to convey their views and conditions all over them.

- Together with the large variety of devices sharing their input on twitter, there is lot of tweet that we need to review to manually collect the grievance information from the resident, which is not an easy task to do.

- The primary factor in the study of sentiment is:
  - Data collection.
  - Pre-processing.
  - Extraction of features.
  - Choice of features.
  - Classification.

Case study

- Content data can be described as a new field of computing technology study to address the lack of knowledge through:
  - Integration of information processing.
  - Artificial intelligence.
  - Data extraction.
  - NLP.
  - Content development strategies.

- Document extraction is a portion of the document prospecting field reliant on searching for information trends appropriate to the level of large code files.

Step 1: Document collection

Keyword	Output
‘flood', 'trash', `education`, `service;	9 localities in the city of Surabaya have the potential to occur flooding after being rained in overnight http: //t.co/.
	Tear floods occur during the funeral process of community leaders in Surabaya

Pre possessing

- Case folding.
- Tokenizing.
- Stemming.
- Stop word removal.

Case folding

Input	Output
9 localities in the city of Surabaya have the potential to occur flooding after being rained in overnight http: //t.co/.	localities in the city of Surabaya have the potential to occur flooding after being rained in overnight.
Tear floods occur during the funeral process of community leaders in Surabaya	tear floods occur during the funeral process of community leaders in surabaya

Tokenizing

Input	Output
localities in the city of Surabaya have the potential to occur flooding after being rained in overnight.	[‘localities’,‘in’,‘the’,‘city’,‘of’,‘Surabaya’, ‘have’, ‘the’, ‘potential’, ‘to’, ‘occur’, ‘flooding’,‘after’, ‘being’, ‘rained’, ‘in’, ‘overnight’,.
tear floods occur during the funeral process of community leaders in surabaya	[‘tear’, ‘floods’, ‘occur’, ‘during’, ‘the’, ‘funeral’, ‘process’, ‘of’, ‘community’, ‘leaders’, ‘in’, ‘surabaya’,]

Stemming

Input	Output
localities in the city of Surabaya have the potential to occur flooding after being rained in overnight	[‘local’,‘in’,‘the’,‘city’,‘of’,‘Surabaya’, ‘have’, ‘the’, ‘potential’, ‘to’, ‘occur’, ‘flood’,‘after’, ‘be’, ‘rain’, ‘in’, ‘overnight’,]
tear floods occur during the funeral process of community leaders in surabaya	[‘tear’, ‘floods’, ‘occur’, ‘due’, ‘the’, ‘funeral’, ‘process’, ‘of’, ‘community’, ‘leader’, ‘in’, ‘surabaya’,]

Stop word removal

Input	Output
[‘local’,‘in’,‘the’,‘city’,‘of’,‘Surabaya’, ‘have’, ‘the’, ‘potential’, ‘to’, ‘occur’, ‘flood’,‘after’, ‘be’, ‘rain’, ‘in’, ‘overnight’,.]	[‘local’,‘city’,,‘Surabaya’, ‘potential’, ‘occur’, ‘flood’, ‘rain’, ‘overnight’,]
[‘tear’, ‘flood’, ‘occur’, ‘due’, ‘the’, ‘funeral’, ‘process’, ‘of’, ‘community’, ‘leader’, ‘in’, ‘surabaya’,]	[‘tear’, ‘flood’, ‘occur’, ‘due’, ‘funeral’, ‘process’, ‘community’, ‘leader’, ‘ surabaya’,]

Step 3: Feature extraction

- Feature extraction is a method of turning information into a set of characteristics to gain information related to the objectives of gathering minimal representation of data.

- Token count vectorizer:

Input	Vocabulary
['local','city',,'Surabaya', 'potential', 'occur', 'flood', 'rain', 'overnight',]	['local','city',,'Surabaya', 'potential', 'occur', 'flood', 'rain', 'overnight','tear', 'flood', 'occur', 'due', 'funeral', 'process', 'community', 'leader', ' surabaya',]
['tear', 'flood', 'occur', 'due', 'funeral', 'process', 'community', 'leader', ' surabaya',]	
Vocabulary	Output
['local','city',,'Surabaya', 'potential', 'occur', 'flood', 'rain', 'overnight','tear', 'flood', 'occur', 'due', 'funeral', 'process', 'community', 'leader', ' surabaya',]	[0=>1, 1=>1, 2=>1, 3=>1, 4=>1, 5=>1, 6=>1, 7=>0, 8=>0, 9=>0, 10=>0, 11=>0, 12=>0]
	[0=>0, 1=>0, 2=>0, 3=>0, 4=>1, 5=>0, 6=>0, 7=>1, 8=>1, 9=>1, 10=>1, 11=>1, 12=>1]

Term frequency-Inverse document frequency

- The TF-IDF method is a tool used to measure the meaning of each term normally utilized for data retrieval.

- For each token in each corpus document, this approach will calculate the values of Term Frequency (TF) and Inverse Document Frequency (IDF) with the equation shown in equation 1.

- Equation 1: $$W(dt) = tf(dt) * idf(t)$$

- Equation 2: $$idf(t) = \log(D/df)$$

Step 4: Feature selection

- The selection of features aims to improve the efficiency of the classification process by lowering the amount of tokens evaluated, so the research that will be done will be just a few.

- In this case study, we use the PHP programming language machine learning library which facilitates choice of functions.

- Variance threshold:
  - Select the factor amount from other factor to reflect the characteristic we will are using in the identification system by choosing token that is above an information variability value-based (default:0).

Step 4: Classification

- A variety of controlled learning methods we can utilize are:
  - Naïve Bayes classifier.
  - Confusion matrix.
  - Labelling results.

How to implement

- Two parts:
  - Section education.
  - Section evaluation,

- System design will be designed.

- The program will check for tweets corresponding to the keywords mentioned.

- Once information is accessible, the post obtained from the prior phase will be analysed on attempt such as the desire we address in chapter 3 procedures before we find some categorization for the post.

- The programs will use the labelling provided by the administrator for more training and on the following identification phase as an information, so devices will have more reliable forecasts when the systems detect a similar tweet.

Result

- In this study, 3000 tweet datasets were largely used from input from Pemerintah Kota Surabaya.

- We are conducting tests on our first phase post identification method at this stage. The testing was carried out spontaneously using the separated method of the train test, where the assessment process was separated into two.

Test Number 1					
Predict					
Positive Negative					
	Actual	Positive			
Negative		TP:49		FP: 6	
FN:12	TP:33				

Checkpoint (1 of 2)

Multiple choice questions:

1. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?
   a) Decision tree
   b) Regression
   c) Classification
   d) random forest

2. To find the minimum or the maximum of a function, we set the gradient to zero because:
   a) The value of the gradient at extrema of a function is always zero
   b) Depends on the type of problem
   c) Both A and B
   d) None of the above

3. The most widely used metrics and tools to assess a classification model are:
   a) Confusion matrix
   b) Cost–sensitive accuracy
   c) Area under the ROC curve
   d) All the above

Checkpoint solutions (1 of 2)

Multiple choice questions:

1. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?
   a) Decision tree
   b) Regression
   c) Classification
   d) Random forest

2. To find the minimum or the maximum of a function, we set the gradient to zero because:
   a) The value of the gradient at extrema of a function is always zero
   b) Depends on the type of problem
   c) Both A and B
   d) None of the above

3. The most widely used metrics and tools to assess a classification model are:
   a) Confusion matrix
   b) Cost–sensitive accuracy
   c) Area under the ROC curve
   d) All the above

Checkpoint (2 of 2)

Fill in the blanks:

1. A ---------- is a piece of software that conducts a conversation via auditory or textual methods.
2. **----------** a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector.
3. **----------** is the process of reducing the number of input variables when developing a predictive model.
4. **----------** basically gives us an idea about how well our classifier has performed, with respect to performance on individual classes.

Checkpoint solutions (2 of 2)

Fill in the blanks:

1. A **chatbot** is a piece of software that conducts a conversation via auditory or textual methods.
2. **Feature extraction** a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector.
3. **Feature selection** is the process of reducing the number of input variables when developing a predictive model.
4. **Confusion matrix** basically gives us an idea about how well our classifier has performed, with respect to performance on individual classes.

Question bank

Two mark questions:

1. List AI use cases for citizen inquires and information.
2. Define data management services.
3. Define media richness theory.
4. Define naive Bayes classifier.

Four mark questions:

1. List some of the AI use cases for citizen inquires and information.
2. How catboats can be used for communication between citizen and government.
3. How machine learning can be used for drafting documents.
4. List the applications of machine learning in solving citizen problems.

Eight mark questions:

1. Write a short note on how machine learning techniques can be used in chat bot for communication between citizen and government.
2. Which machine learning techniques can be used for drafting documents.

Unit summary

Having completed this unit you should be able to:

- Understand the type of government problems appropriate for AI applications
- Learn some citizen service use cases
- Understand the importance of chat boots for communication between citizen and government
- Gain knowledge on applications of machine learning techniques to understand the sentiment analysis of citizen
- Learn about machine learning for classification of citizen complaints