

Topics for Capstone 1 & 2

E: mannd@duytan.edu.vn

C1. Android Complaint Management System

The system consists of an android application as client app and a pc acting as web server. The user must have the android app installed on his mobile phone. On first time use, the user may now have to register on the application. The user details are now stored on the server db. Next the user may now login using username and password. After login, the user now has access to the complaint form. User may post a complaint on the application and post it. The complaint message is sent to the server. Server may now reply to the complaint and the reply is sent on users android phone.

Modules:

- **Login**
- **Registration**
- **Complaint posting**
- **Complaint Viewing on server**
- **Complaint reply posting**
- **Complaint reply transfer to android device**

Software Requirement:

- Android Sdk
- Jdk 6.0 or higher
- Windows 7 or higher

Hardware Requirements:

- Processor – i3
- Hard Disk – 5 GB
- Memory – 1GB RAM

- Android device

Reference:

- <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5578355&queryText%3DAndroid+based+shoping>
- <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6040585&queryText%3DAndroid+based+shoping>
- <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5967292&queryText%3DAndroid+and+pc+connection>

C2. An Android Chatbot

An Android Chatbot is build using algorithm that analyzes user's queries and understand user's message. This System provides answers to the queries of the Users. User just have to enter his query and the system analyzes the key words and answers the query. The user can ask questions about the different categories like sports, education, health, etc. The system will compute the question with highly optimize algorithm and respond to the user.

If the answer is correct then user can send feedback by either liking the answer or by the sending feedback and if the answer is incorrect then system will provide option to search on Google. The Admin can view user's feedback and it's like and dislike on different questions. System allows admin to manage all the users and questions.

1. **Admin Login:** Need to enter login credentials into web interface by the admin or the authorized person.
 - **Add/View Questions:** Can add a new question into the system and also can view it along with it's like and dislike.
 - **View Users:** Can view all the users list who are registered with system.
 - **View Feedback:** Can view all the users feedback with there Email Id and date time.
2. **User Login:** Here, user need to enter the login credentials into android app for login and if He/she does not have login credential they need to **Sign Up** first.
 - **View Profile:** Here, user can view their own profile.
 - **Ask Question:** Here, user can ask questions directly with system and if answer is found then user can like or dislike the answer. Or if answer is not found then user can get option to search on Google.
 - **Send Feedback:** Here, User can send their feedback to admin.

Front End: Android

Backend: ASP.net with C#

Hardware Requirement:

- i3 Processor Based Computer
- 1GB-RAM
- 80 GB Hard Disk
- Monitor
- Internet Connection
- Android Device

Software Requirement:

- Windows 7 or higher
- Android Studio
- SQL Server 2008

Advantages:

- Chat bot for answering the user's query in fastest way.
- This application saves time for the users to get all answer at one place.

Disadvantages:

- Requires active internet connection.
- System may provide inaccurate results if data not entered properly.

Applications:

- The system can be used for schools, college, or universities as a smart assistance.

C3. Student Information Chat Bot System

User interfaces for software applications can come in a variety of formats, ranging from command-line, graphical, web application, and even voice. While the most popular user interfaces include graphical and web-based applications, occasionally the need arises for an alternative interface. Whether due to multi-threaded complexity, concurrent connectivity, or details surrounding execution of the service, a chat bot based interface may suit the need.

Chat bots typically provide a text-based user interface, allowing the user to type commands and receive text as well as text to speech response. Chat bots are usually a stateful services, remembering previous commands (and perhaps even conversation) in order to provide functionality. When chat bot technology is integrated with popular web services it can be utilized securely by an even larger audience.

- A Student Information Chat Bot project is built using artificial algorithms that analyzes user's queries and understand user's message.
- This System is a web application which provides answer to the query of the student very effectively.
- Students just have to query through the bot which is used for chatting.
- Students can chat using any format, as there is no specific format that the user has to follow.
- The System uses built in artificial intelligence to answer the query.
- The answers are appropriate what the user queries.
- If the answer found to be invalid, user just need to select the invalid answer button which will notify the admin about the incorrect answer.
- Admin can view invalid answer through portal via login

- System allows admin to delete the invalid answer or to add a specific answer of that equivalent question.
- The user does not have to personally go to the college for enquiry.
- The system analyzes the question and then answers to the user.
- The system replies using an effective Graphical user interface which implies that as if a real person is talking to the user.
- The user can query about the college related activities through online with the help of this web application.
- This system helps the student to be updated about the college activities.

Features:

The system comprises of 3 modules as follows:

- 1. Admin Login**
- 2. Bot Chat**
- 3. Text to Speech**

Description:

1. Admin Login:

- User has to login to the system to access various helping pages through which user can ask queries to the system with the help of bot.

2. Bot Chat:

- User can chat with the bot it implies as if enquiring to the college person about college related activities.

3. Text to Speech:

- The bot also speaks out the answer.

Software Requirements:

- Windows 7 or higher

- MySql
- Wamp Server

Hardware Components:

- Processor – i3
- Hard Disk – 5 GB
- Memory – 1GB RAM

Advantages

- User does not have to go personally to college office for the enquiry.
- This application enables the students to be updated with college cultural activities.
- This application saves time for the student as well as teaching and non teaching staffs.

Application

Enhance AI Based Net Student System can be used in many colleges around the country and it can be used in various firms.

C4. Virtual Personal Assistant Apps development

Problem Statement

Artificial Intelligence personal assistants have become plentiful over the last few years. Applications such as Siri, Bixby, Ok Google and Cortana make mobile device users' daily routines that much easier. You may be asking yourself how these functions. Well, the assistants receive external data (such as movement, voice, light, GPS readings, visually defined markers, etc.) via the hardware's sensors for further processing - and take it from there to function accordingly.

Not too long ago, building an AI assistant was a small component of developers' capacities; however, nowadays, it is quite a realistic objective even for novice programmers. To create a simple personal AI assistant, one simply needs dedicated software and around an hour of working time. It would take much more time, though, to create something more advanced and conceptually innovative. Nonetheless, well thought-out concepts can result in a great base for a profitable startup. Let us consider the six most renowned applications based on artificial intelligence concepts that can help create your virtual AI assistant app.

Background

Siri. Siri is Apple Inc.'s cloud software that can answer users' various questions and give recommendations, due to its voice processing mechanisms. When in use, Siri studies the user's preferences (like contextual advertising) to provide each person with an entirely individual approach. This software solution is also useful for developers; the presence of API called SiriKit provides smooth integration with new applications developed for iOS and watchOS platforms.

Ok Google. Ok Google is an Android-based voice recognition application, which is launched by users uttering commands of the same name. This software features very advanced functions including web search, route optimization, memo scheduling etc. that can collectively help users solve a wide array of daily tasks. Like Siri, the creators of Ok Google offer Google Voice Interaction API. This interface can become a truly indispensable tool in the development of mobile applications for the Android platform.

Cortana. A virtual intelligent assistant with the function of voice recognition and AI elements, Cortana was developed for such platforms as Windows, iOS, Android, and Xbox One. It can predict users' wants and needs based on their search requests,

e-mails, etc. One of Cortana's distinguishable features is her sense of humor. "She" can sing, make jokes and speak to users informally.

Amazon Echo. Amazon Echo combines hardware and software that can search the web, help with scheduling of upcoming tasks and play various sound files all based on voice recognition. A small speaker equipped with sound sensors, the device can be automatically activated by exclaiming "Alex."

Nina. Software with AI elements that has a main goal of narrowing down the amount of physical effort spent on the solution of daily tasks (web search, scheduling, etc.) Due to elaborate analytical mechanisms, Nina becomes "smarter" with every day of personal utilization.

Bixby. Samsung's Bixby application is another successful implementation of the AI concept. It also builds a unique user approach, based on interests and habits. Bixby features advanced voice recognition mechanisms, and uses the camera to identify images, based on markers and GPS.

Methodology

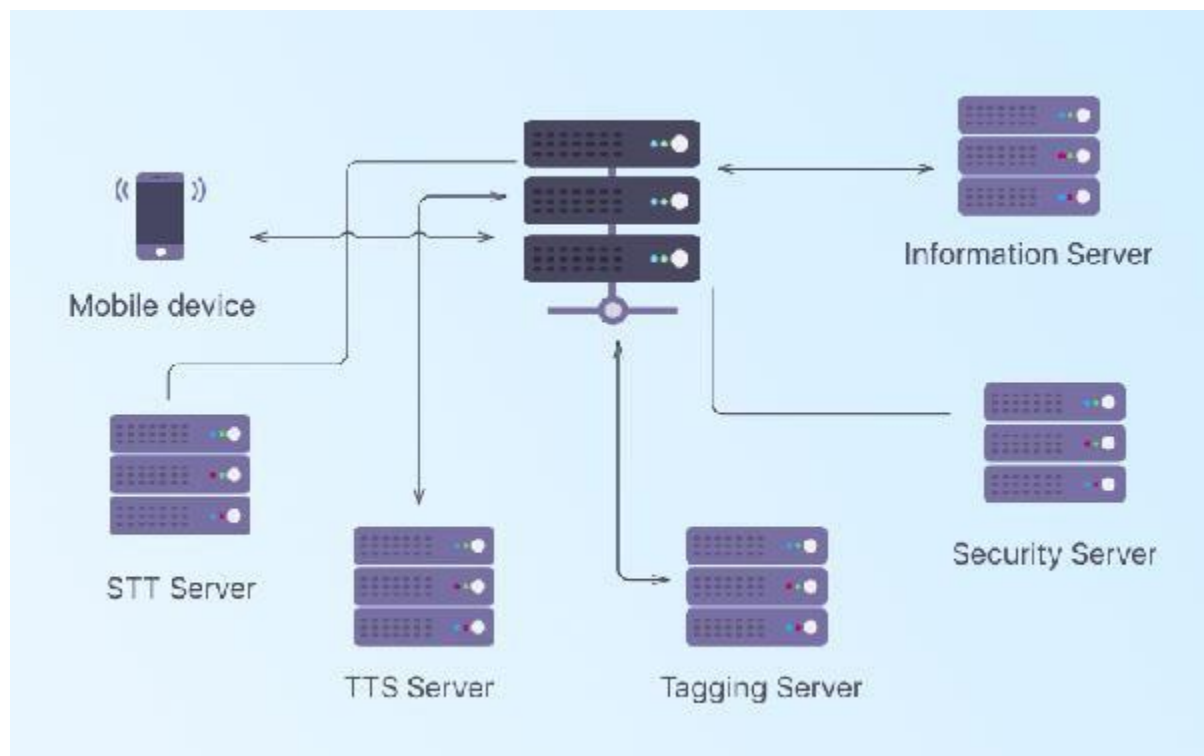


Fig: 1 Mobile voice assistant's architecture

The general operating principle of artificial intelligence assistants is the ability to make personal decisions based on incoming data. The software has to include an advanced set of tools for processing received data, in order to make proper individual choices. Artificial neural networks were invented to help develop the discussed software. Such networks imitate the human brain's ability to remember, to help the assistant recognize and classify data and customize predicting mechanisms based on thorough analysis. The memory process is executed deductively, i.e. top-down: first, the app analyzes several variants of outcome; then, it remembers the variants applied by a human (i.e. the system remembers proper answers to the question "How are you?" such as "I'm fine", "Not very well" etc., and ignores answers like "Yes", "No" and others) and "self-educates" to be able to generate situation-based algorithms later. It is not necessary to manually enter information into the app to build your own personal artificial intelligence assistant. API software was developed for that, and the application programming interface aids the apps in the recognition of faces, speech, documents and other external factors. There are a number of APIs on the market, most popular of which are api.ai, Wit.ai, Melissa, Clarifai, Tensorflow, Amazon AI, IBM Watson; with less widespread options including Cogito, DataSift, iSpeech, Microsoft Project Oxford, Mozscape and OpenCalais. Let us examine some of these.

Experimental Design

How to Create Virtual Assistant Apps like Siri and Google

Assistant Developing your own voice assistant app

If you intend to make your own Siri or Google assistant, make sure that you do possess the appropriate skills and sources, because this process is far from simple. Basic technologies in mobile voice assistants.

Voice/speech to text (STT)

This is the process of converting speech signal into digital data (e.g., text data). The voice may come as a file or a stream. You can use CMU Sphinx for its processing.

Text to speech (TTS)

This is the opposite process that translates text / images in a human speech. It is very useful when, for instance, a user wants to hear the correct pronunciation of a foreign word.

Intelligent tagging and decision making

Intelligent tagging and decision making serve for interpreting the user's request. For example, the user may ask: 'What do I watch tonight?'. The technology will tag the top-rated movies and suggest you a few according to your interests. The AlchemyAPI may help you in the implementation of this task.

Image recognition

Image recognition is an optional but very useful picture. Later, you can use it for developing multimodal speech recognition. Have a look at OpenCV if you are thinking of developing it.

Noise control

The noises from cars, electrical appliances, other people talking near you make the user's voice unclear. This technology will reduce or eliminate the background noise that prevents a correct voice recognition.

Voice Biometrics

This is a very important option from the point of view of security. Thanks to this feature, the voice assistant may identify who is talking and whether it is necessary to respond. Thus, you may avoid a comic situation that happened to Siri and Amazon Alexa when they lowered the temperature in a house and even turned off someone's thermostat by hearing a relevant command from the TV speakers.

Speech compression

With this mechanism, the client side of the applications will resize the voice data and send it to the server in a succinct format. It will provide a fast application performance without annoying delays.

Voice interface

Voice interface is what the user hears and sees in return to his or her request. For the voice part, you will need to pick up the voice itself, set the rate of speech, the manner of speaking, etc. For the visual part, you will have to decide on the visual representation that a user is going to see on the screen. If reasonable, you can skip it at all. Note that voice and text data may be processed either on a server or directly

within a device. In the picture below, we have shown the scheme that works with the server participation.

References

1. https://www.researchgate.net/publication/264001644_Virtual_Personal_Assistant
2. <https://artjoker.net/blog/how-to-use-artificial-intelligence-in-mobile-apps/>
3. <https://www.brainasoft.com/brainasoft/>

C5. Question answering system for automated customer relationship management

Problem Statement

The problem of automated question answering system is to automatically answer the questions without waiting for some personnel to handle the query. Question answering (QA) is a well- researched problem in NLP. In spite of being one of the oldest research areas, QA has application in a wide variety of tasks. Recently, QA has also been used to develop dialog systems and chatbots designed to simulate human conversation. However, with recent developments in deep learning, neural network models have shown promise for QA. Although these systems generally involve a smaller learning pipeline, they require a significant amount of training. GRU and LSTM units allow recurrent neural networks (RNNs) to handle the longer texts required for QA. Further improvements – such as attention mechanisms and memory networks – allow the network to focus on the most relevant facts. Such networks provide the current state-of-the-art performance for deep-learning-based QA. The challenges which are faced while designing the Automated QA CRM is how to deal with individual human ,how to understand his/her sentences if they are misspelled/ shortforms , how much time it needs to understand a human, how well it map with the intent of human.

Background Work

Despite of the challenges given in the problem statement self driving car is still an active area of research. Numerous approaches have been proposed over the years. In spite of being one of the oldest research areas, QA has application in a wide variety of tasks, such as information retrieval and entity extraction. Recently, QA has also been used to develop dialog systems and chatbots designed to simulate human conversation. Traditionally, most of the research in this domain used a pipeline of conventional linguistically-based NLP techniques, such as parsing, part-of-speech tagging and co-reference resolution. Many of the state-of-the-art QA systems – for example, IBM Watson use these methods.

Prior to this work there has been application of various concepts like attention mechanisms to the CNN. There is another important concept of memory network, dynamic memory networks. Other researcher applies end-to-end memory networks to achieve state-of-the-art results with weak supervision. Other researcher also focused on a template-based approach to SPARQL query generation and handle constructs that are not captured using semantic triple representation

Methodology

1. **Data Collection and Dataset Preparation:** This will involve collection of data from various sources. In this research the data is collected as 2 dataset of MCTest.
2. **Training:** *The data is tested and trained using a baseline model built with keras and tensorflow –a GRU model*
3. **Deployment and analysis on real life scenario:** The model developed is deployed for further analysis where both positive and negative cases will be used for further improvement in methodology.

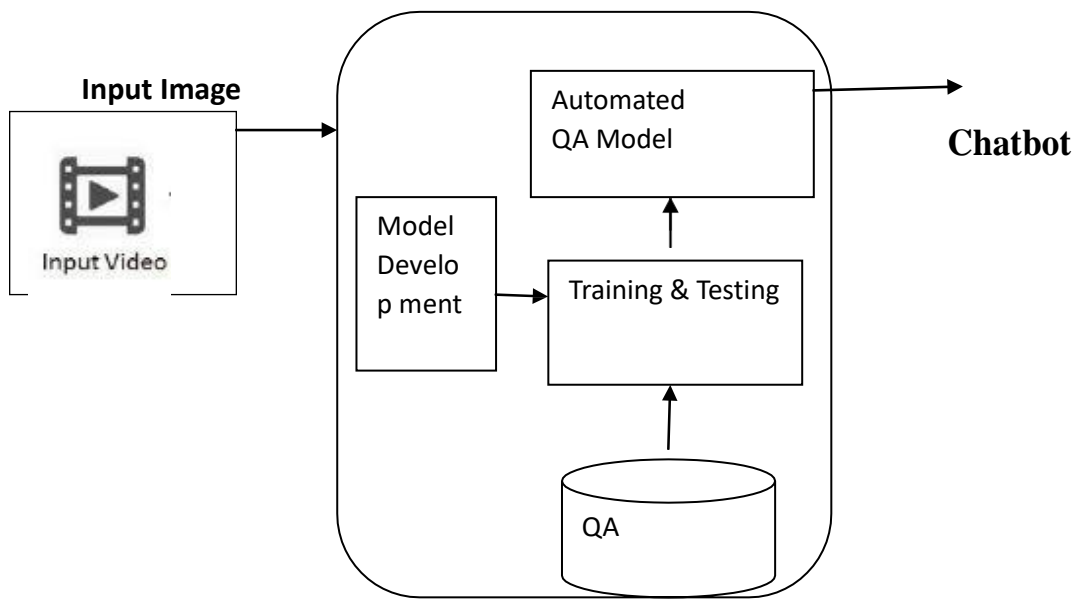


Figure 1: Architecture of Automated QA CRM System

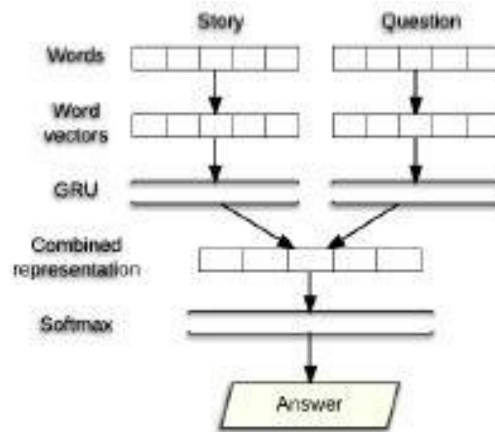


Figure 2:GRU Baseline used as model [Eylon Stroh et al. Question Answering using Deep learning]

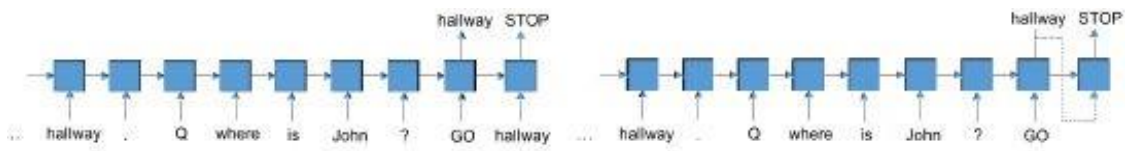


Figure 3: Seq-to-Seq Baseline used as model [Eylon Stroh et al. Question Answering using Deep learning]

Experimental Design

Dataset: Automated QA dataset is made by gathering two MCTest dataset.

Evaluation Measures: Measures such as accuracy and time to respond will be used.

Software and Hardware Requirements: *Python* programming language, Keras, tensorflow and NVIDIA GTX 1070 GPU 8 GB.

C6. Predicting Student Performance with Deep Neural Networks

Problem Statement

In present educational systems, student performance prediction is getting worsen day by day. Predicting student performance in advance can help students and their teacher to keep track of progress of a student. Many institutes have adopted continuous evaluation system today. Such systems are beneficial to the students in improving performance of a student. The purpose of continuous evaluation system is to help regular students.

In recent years, Neural Networks have seen widespread and successful implementations in a wide range of data mining applications, often surpassing other classifiers. This study aims to investigate if Neural Networks are a fitting classifier to predict student performance from Learning Management System data in the context of Educational Data Mining. To assess the applicability of Neural Networks, we compare their predictive performance against six other classifiers on this dataset. These classifiers are Naive Bayes, k-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine and Logistic Regression and will be trained on data obtained during each course. The features used for training originate from LMS data obtained during the length of each course, and range from usage data like time spent on each course page, to grades obtained for course assignments and quizzes. After training, the Neural Network outperforms all six classifiers in terms of accuracy and is on par with the best classifiers in terms of recall. We can conclude that Neural Networks outperform the six other algorithms tested on this dataset and could be successfully used to predict student performance.

Background

In recent years, the use of internet-based educational tools has grown rapidly (Jordan, 2014) as well as the research surrounding them (see Figure 1). These tools provide a clear advantage for students and teachers alike, with the ability to access and share course data from anywhere in the world, track student progress and provide rich educational content.

Amrieh, et al. proposed a prediction model for students' performance based on data mining methods with some few features called student's behavioral features.

The model was evaluated in three different classifiers; Naïve Bayesian, Artificial Neural Network and Decision tree. Random Forest, Bagging and Boosting were used as ensemble methods to improve the classifier's performance. The model achieved up to 22.1% more in accuracy compared when behavioral features were removed. It increased up to 25.8% accuracy after using the ensemble methods.

The family of classifiers this study focuses on, Neural networks, have shown promising results in domains like speech recognition (Graves & Jaitly, 2014), computer vision (Venugopalan et al., 2014), recognizing music (Costa, Oliveira, & Silla, 2017), playing complex games like GO (Wang et al., 2016) and economic forecasting (Nametala, Pimenta, Pereira, & Carrano, 2016), but their use in EDM has thus far been limited compared to other classification algorithms (Baker & Inventado, 2014). This can be partly explained by their difficulty to set up, the lack of convenient all in one packages that are easy to use and the often-long training times (Gaur, 2012). But they do offer clear benefits over other machine learning algorithms. They can classify instances in domains that are not linearly-separable and can handle noisy and complex data (Schmidhuber, 2015). These properties make them especially suited for a domain like EDM where the data, given the fact that it is based on human behavior, can be complex, might contain irrelevant entries as well as nonlinear relations.

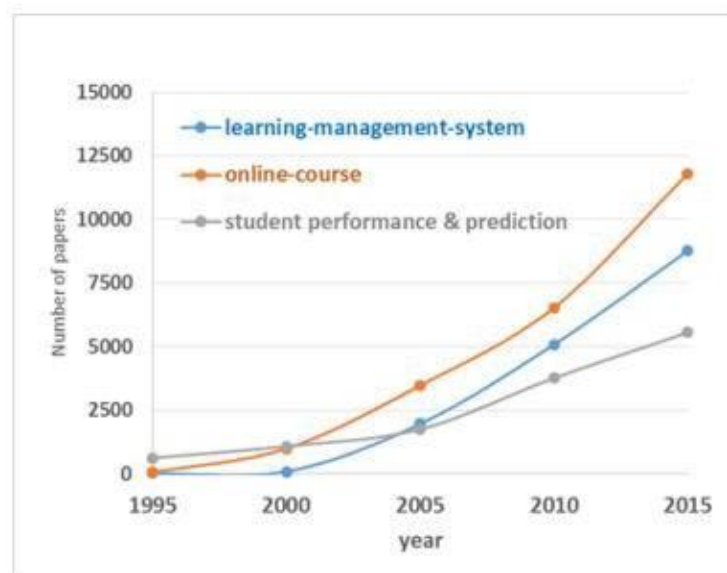


Figure 1: Number of papers in Educational Data Mining related fields. source: Google Scholar.

Methodology

Neural Networks in Student Performance Prediction

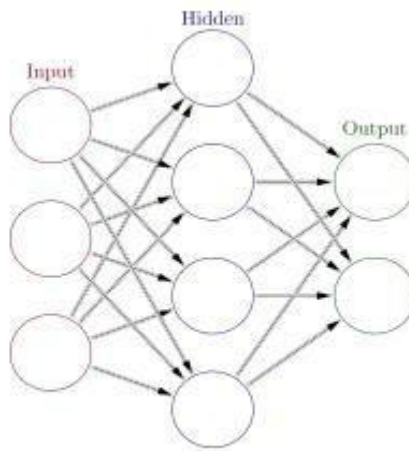


Figure 2: The structure of a Neural Network (Glosser.ca, 2013).

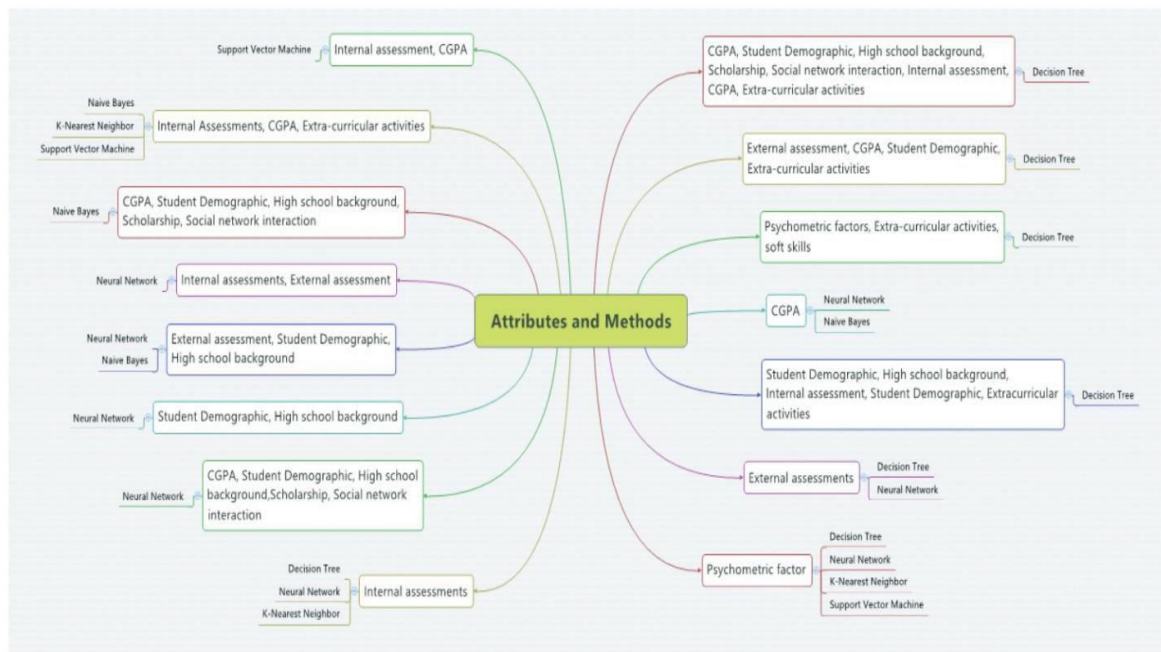


Fig 3: List of common attributes and methods used in predicting student's performance [2]

This study focuses on to predict student performance belongs to the family of Neural Networks. Neural Networks are algorithms that mimic the way our brain

works. They consist of an array of interconnected nodes that exchange information among each other (see Figure 2), comparable to the way our neurons, connected by dendrites and axons, exchange information. They learn iteratively over time by observing different examples, similarly to how children can learn skills from their parents by observation. However, unlike children that can learn to recognize an object after only observing it once, Neural Networks often require a greater set of observations to attain sufficient predictive capacity. In this section the steps that were performed to extract the predictors are explained and the training process and parameter tuning for the classifiers are clarified: -

Dataset:

Link: <https://www.kaggle.com/aljarah/xAPI-Edu-Data>

Name: Students' Academic Performance Dataset

Dataset attributes: -

Students Dataset		
Name	Data Type	Distinct Values
Gender	Nominal	2
Nationality	Nominal	14
Place of Birth	Nominal	14
Stages	Nominal	3
Grades	Nominal	12
SectionID	Nominal	3
Topic	Nominal	12
ParentResponsible	Nominal	2
Semester	Nominal	2
Raised hand	Numeric	0-100
Visited Resource	Numeric	0-100
Viewing Announcement	Numeric	0-100
Discussion Group	Numeric	0-100
Parent Answering	Nominal	2
Parent Satisfaction	Nominal	2
Student Absent day	Nominal	2

Experimental Design

In this problem, we have to build a Deep Neural Network linear classifier model to predict the performance of students. This process should be followed once the dataset is preprocessed: data cleaning and data transformation. The DNN model will be built using python3 and tensorflow 1.3.0.

Artificial neural networks (ANNs) are parallel computational models comprised of densely interconnected, adaptive processing units, characterized by an inherent propensity for learning from experience and discovering new knowledge. Due to their excellent capability of self-learning and self-adapting, they have been extensively studied and have been successfully utilized to tackle difficult real-world problems (Bishop 1995; Haykin 1999) and are often found to be more efficient and more accurate than other classification techniques (Lerner et al. 1999). Classification with a neural network takes place in two distinct phases. First, the network is trained on a set of paired data to determine the inputoutput mapping. The weights of the connections between neurons are then fixed and the network is used to determine the classifications of a new set of data. Although many different models of ANNs have been proposed, the feedforward neural networks (FNNs) are the most common and widely used in a variety of applications.

Python is a full featured for general purpose programming language. It is a mature and fast expanding platform for scientific research and numerical computing. Python host numerous opensource libraries and almost all general-purpose libraries for machine learning which can be further use for deep learning models. All this benefits from the python ecosystem lead to the top two libraries for numerical analysis of deep learning was developed for python language, that is Tensorflow and Theano library.

TensorFlow is an open source library for computing numerical using data flow graphs. The data flow graphs are also known as Static Computation graph. A developer must first design the input layer and connect every input layer to the hidden layer then the same from hidden layer to output layer. The graphs are made of tensors and ops, defining all the neural networks and all mathematical calculations. The session helps to run the graph. Tensorflow comes with Graphical Processing Unit package where all the matrix calculations can be done efficiently and much faster. Once data is preprocessed, the data is divided into two parts training and testing dataset. It is divided in the ratio 3:1 (Train/Test). In training dataset, the features and classes are split and stored in a tensorflow placeholder. Both training dataset classes records are One-hot encoded, it is a process where

class variables are converted into a numerical form that will be provided to deep neural network model for effective prediction.

Other approaches: -

- Network-Based Clustering
- Baseline Methods

References

1.

[http://delivery.acm.org/10.1145/640000/638180/s2.pdf?ip=115.249.53.74
&id=638180&
cc=OPEN&key=045416EF4DDA69D9%2E0B5B0253775BA20F%2E4D
4702B0C3E38B35%2E6D218144511F3437&_____acm
=1525156531_876a762e556fd775eba267a77d75
e7e0](http://delivery.acm.org/10.1145/640000/638180/s2.pdf?ip=115.249.53.74&id=638180&cc=OPEN&key=045416EF4DDA69D9%2E0B5B0253775BA20F%2E4D4702B0C3E38B35%2E6D218144511F3437&_____acm=1525156531_876a762e556fd775eba267a77d75e7e0)

2. [https://ac.els-cdn.com/S1877050915036182/1-s2.0-
S1877050915036182_main.pdf?_tid=44012f66-b408-473b-
a41b-
69790c3b51e2&acdnat=1525166323_ac1e1a0715d091a4ed3f3
8939666bfd4](https://ac.els-cdn.com/S1877050915036182/1-s2.0-S1877050915036182_main.pdf?_tid=44012f66-b408-473b-a41b-69790c3b51e2&acdnat=1525166323_ac1e1a0715d091a4ed3f38939666bfd4)

C7. Personalized marketing and targeted advertising

Problem Statement

The rapid pace of innovation in the field of Artificial Intelligence (AI) is turning the far-fetched dreams of yesterday into advances of today. Artificial intelligence matches our desires with the exact information, product or service we need, at the right time. It is increasingly being used to improve search results, and it will be the driving force behind the changes in content production, target marketing and advertising.

Personalization based on data has become critically important in digital marketing [1], but many marketers still use outdated methods, with business rules created through online analysis of collected data.

These rules are implemented in an automated way via a business rules engine the accumulates all the rules over time, including old efforts that no longer work. This process is very inefficient for driving customer acquisition, conversion and retention effectively.

Background

Personalized marketing with machine learning [1][2]



The approach for personalized marketing used by the most successful digital companies (Google, Netflix, Instagram, Twitter) is machine learning. Self-

learning sense-and-respond systems do the challenging work of analyzing data and writing business rules in real time. Once these rules are applied, the data analytics system evaluates the efficacy of the rules based on success or failure, and adjusts accordingly. This way, the rules evolve and improve over time, continually advancing business outcomes by delivering insights for marketers.

Advertising using Artificial Intelligence and Deep Learning

<https://www.rtbhouse.com/deep-learning-leads-to-ultra-precise-personalization/>

Deep learning is the next major area of AI-based research, and it will bring a new era of marketing, which both advertisers and end-users will benefit from.

For example, Facebook has an abundance of data from its users which, when mined intelligently, can uncover hidden patterns. Such information can be implemented properly into content creation, marketing strategy or advertising targeting to give users the most efficient information, tailored and customized through AI.

Furthermore, a recent announcement by Coca-Cola indicates that they want to use AI bots to create music for ads, write scripts, post on social media, and buy media – implying that the deep learning ads' revolution seems closer than ever.

Methodology

Personalized marketing is dependent on technology for data collection, data classification, data analysis, data transfer, and data scalability. Technology enables marketing professionals to collect first-party data such as gender, age group, location, and income and connect them with third-party data like click-through rates of online banner ads and social media participation.

Data Management Platforms

Approach to collecting and activating data

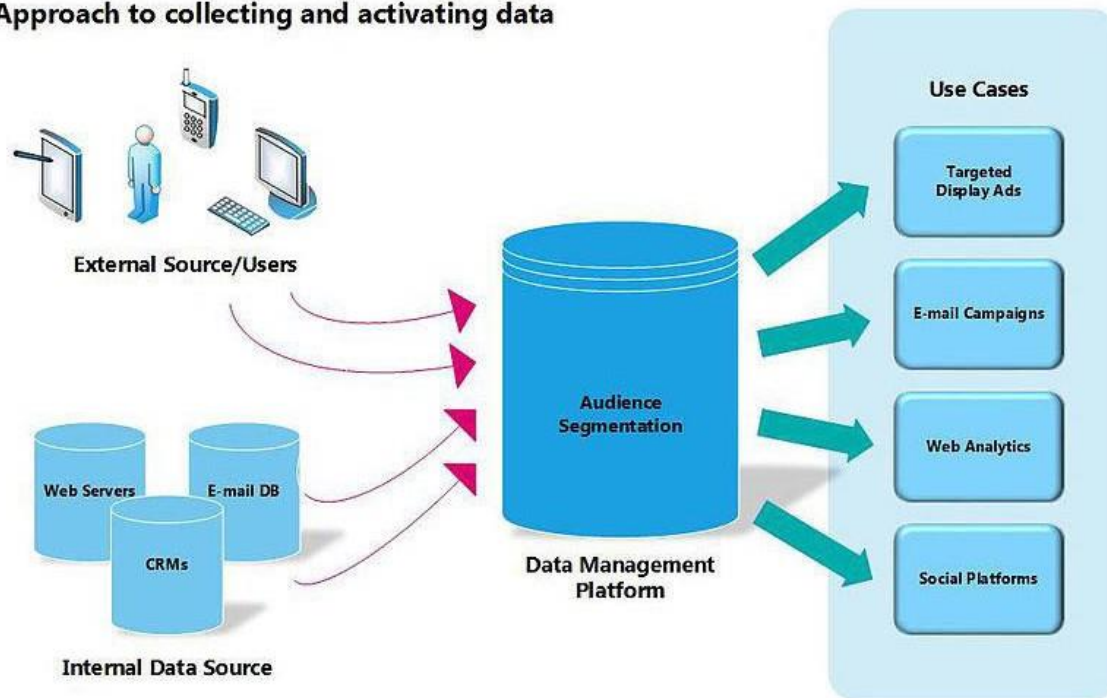


Fig1: Example of data collection[3]

Having knowledge of the consumer's preferences enables suggesting specific products and promotions to each consumer. One-to-one marketing is based in four main steps in order to fulfill its goals:

Identify: To collect reliable data about the preferences of the customers of the company.

Differentiate: To distinguish the customers in terms of their lifetime value to the company, their priorities and needs and to segment them into more restricted groups.

Interact: To get the customer's attention by engaging with him through the communication channels and in ways that he enjoys the most.

Customize: To personalize the product or service to the customer individually using the knowledge that a company has about a customer via machine learning and deep learning tools.

Experimental Design

Step 1: Data collection and dataset preparation

The data set can be generated from specific company or downloaded from www.kaggle.com.

Step 2: Developing a recommender system based on predictions

A structured DNN will be trained for predictive analytics. Also, feature design and selection will be done using various machine learning approaches like Linear Regression, SVM, Random Forest, k-Nearest Neighbors (kNN) etc.

Step 3: Training and experimentation on datasets

The recommender model will be trained for the chosen dataset.

Step 4: Deployment and analysis on real life scenario

The trained and tested recommender system will be developed in real-life scenario.

Experimental Design

Dataset

The dataset from some company will be used for experimentation.

Evaluation measures

Measures such as accuracy will be computed by comparing the desired satisfaction for level of marketing and advertisement.

Software and Hardware Requirements

Deep learning libraries will be exploited for the development and experimentation of the project. Training will be conducted on NVIDIA GPUs for training the DNN model.

References

1. https://scholar.google.co.in/scholar?q=Personalized+marketing+and+targeted+advertising+using+machine+learning&hl=en&as_sdt=0&as_vis=1&oi=scholart&sa=X&ved=0ahUKEwixx6eQ89naAhVMtI8KHd0QA2QQgQMILjAA
2. <https://arabianmarketer.ae/machine-learning-will-transform-creative-content-production/>
3. <https://www.adobe.com/insights/using-machine-learning-enhanced-marketing-results.html>

C8. Face Recognition & Expression Recognition

Mobile App for Visually Impaired Person

Problem Statement

Approximately 285 million people worldwide are visually impaired. One of the most difficult tasks faced by the visually impaired is identification of people. Visually impaired people are unaware of dangers in front of them, even in familiar environments, in unfamiliar environments; such people require guidance to reduce the risk of bumping with obstacles. Voice recognition is a common method of identification; it is an intuitive and difficult process. The rise of computation capability of mobile devices gives motivation to develop applications that can assist visually impaired persons and give them a better life. In this research, the design and implementation of a face detection and recognition system for the visually impaired through the use of mobile computing and deep learning is proposed. Also, a wearable face recognition system for individuals with visual impairments is proposed. This mobile system is assisted by a server-based support system. The challenges of the system lie in better recognition techniques for difficult situations in terms of lighting and weather. Due to the limitation of energy on mobile devices, implementation of a face detection and recognition system based on convolutional neural networks (CNN) that provides detection and recognition services to mobile devices with low hardware specifications is proposed.

Background

Automated face recognition has been the focus of extensive research for the past four decades. The approaches for this task can be broadly divided into two categories: 1) Feature-based methods which first process the input image to extract distinctive facial features, such as the eyes, mouth, nose, etc., as well as other fiducial marks and then compute the geometric relationships among those facial points, thus, reducing the input facial image to a vector of geometric features. 2) Appearance-based (or holistic) methods, which attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. Several alternate sensing technologies such as RFID, infrared and sonar have also been used either on their own or in conjunction with

computer vision to aid the visually impaired. Object detection is the process of locating objects in a given environment like cars, faces, people, and landmark. This process involves research approaches like detecting a barcode based system, by presence of doors, Appearance based methods use techniques such as edge detection and histograms to perform detection, using spatial histogram features to represent objects. Some examples of features used in object detection and recognition are the BoW model, Haar-like, LBP, SIFT and SURF. Face recognition uses a Linearly Approximated Sparse Representation-based Classification (LASRC) algorithm, Sparse Representation Classification (SRC) algorithm, Principal Component Analysis (PCA), face recognition system using multi resolution feature fusion. Object detection system also needed, also some mobile applications are present on the Apple Store, Google Play and Windows Phone Store.

Methodology

Step 1: Data collection and dataset preparation

This will involve collection of Images captured using a high-resolution camera and those captured using a smartphone, FERT dataset, ImageNet dataset will be used then the input data is processed into a set of features before becoming suitable inputs for ,convolutional neural networks(CNN) based face recognition and expression recognition system.

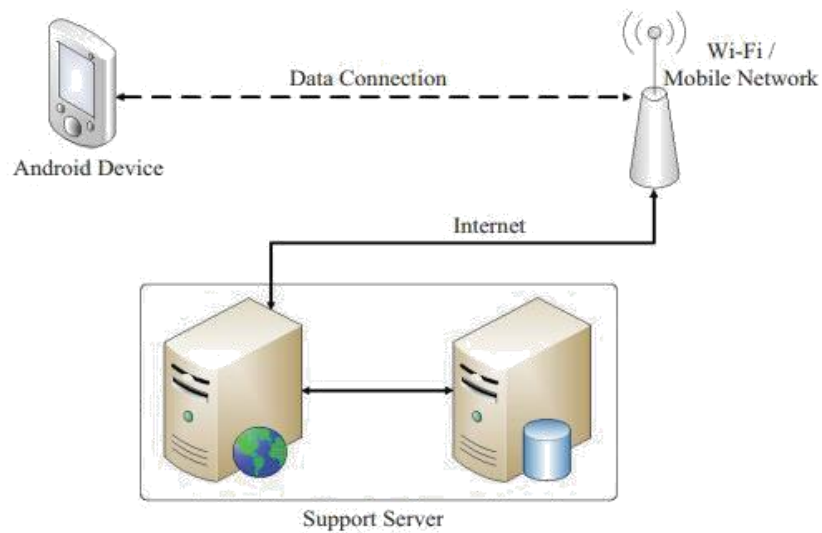
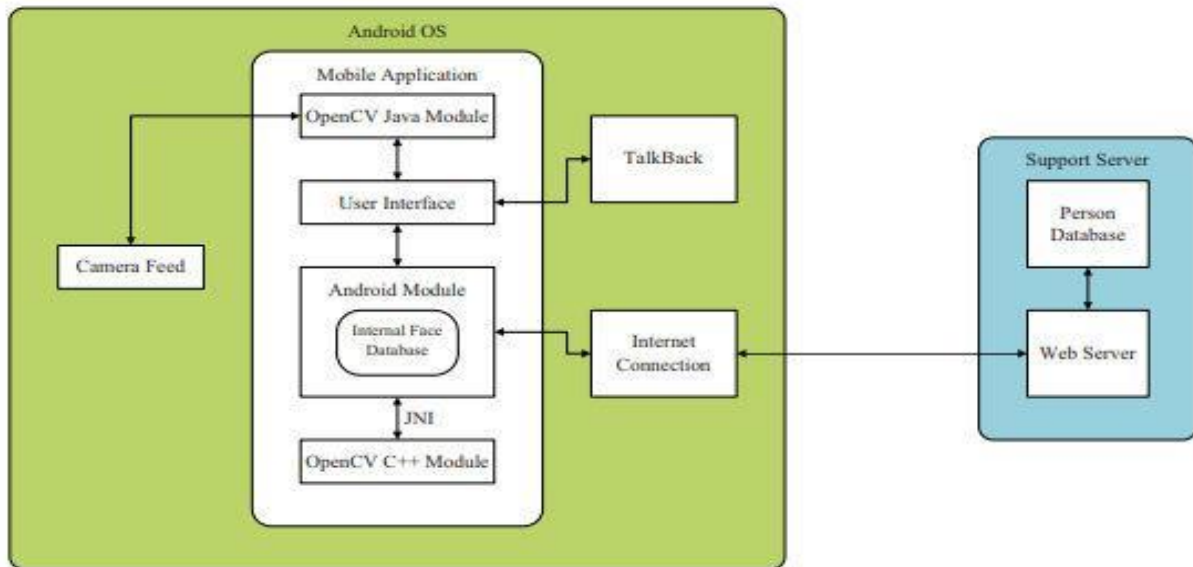
Step 2: Developing a Face Detection and Recognition System for the Visually Impaired Through the Use of Mobile Computing and Wearable Device

In this step Face detection, Face recognition, expression recognition using Cascade Classifier, and convolutional network expression recognition system, android mobile app is developed for the smartphone based system to achieve proposed objective also one wearable device is developed.

Step 3: Training and experimentation on datasets

The convolutional neural networks based face recognition and expression recognition system will be trained on the dataset and images received real time to do face recognition and expression recognition accurately and notify visually impaired persons on time.

Step 4: Deployment and analysis on real life scenario



(b)

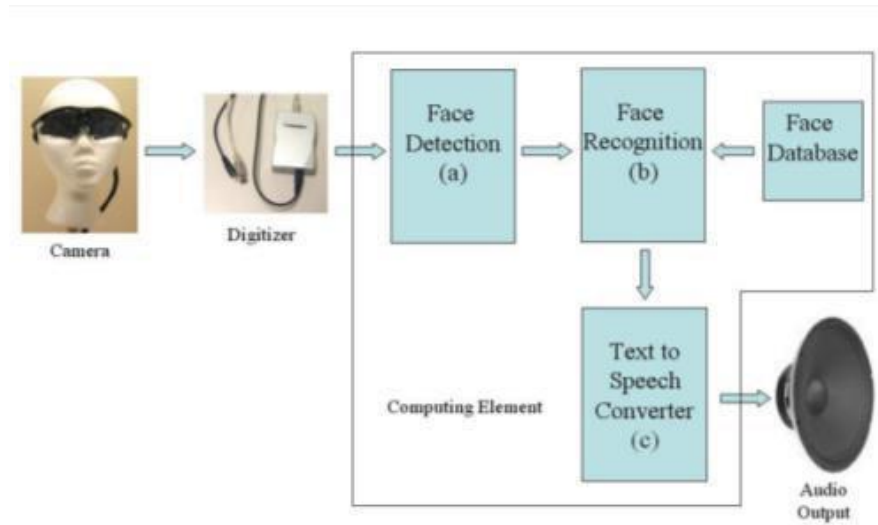


Figure 1 (a) Face recognition system software framework (b) Design of the proposed face recognition & expression recognition mobile app for visually impaired person [Chaudhry, Shonal, and Rohitash Chandra. "Design of a mobile face recognition system for visually impaired persons." arXiv preprint arXiv:1502.00756 (2015).](c) Block diagram of the wearable face recognition system[S. Krishna, G. Little, J. Black, and S. Panchanathan, "A wearable face recognition system for individuals with visual impairments," in Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility, Baltimore, MD, USA, 2005, pp. 106-113]

The trained and tested face recognition and expression recognition model and device will be deployed in a real-life scenario to detect and recognize faces and expressions & will be leveraged for further improvement in the methodology and will follow the above architecture.

Experimental Design

Dataset

FERET dataset (<http://www.face-rec.org/databases/>), ImageNet dataset, Images are captured using a high-resolution camera and those captured using a smartphone will be used for experimentation and evaluation.

Evaluation Measures

Measures such as Mean Average Precision, Battery, Processing Power and Memory, Data Usage, Accuracy, Reliability, Number of recognition of obstacle types, Process Speed, Spend time to transmit image to server are used for face recognition & expression recognition mobile app for visually impaired person.

Software and Hardware Requirements

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python and libraries such as Tensorflow, OpenCV and Keras will be utilized for this process. Android device, Mobile Application, Images using a camera and from a smartphone will be needed. Training will be conducted on NVIDIA GPUs for training the above proposed system that contains smartphone-based guiding system for Face recognition & expression recognition for visually impaired people.

C9. Fake news detection in online social media

Problem Statement

Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media. On the other hand, it enables the wide spread of “fake news”, i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy.

Background

Detecting fake news on social media poses several new and challenging research problems. Though fake news itself is not a new problem—nations or groups have been using the news media to execute propaganda or influence operations for centuries—the rise of web-generated news on social media makes fake news a more powerful force that challenges traditional journalistic norms. There are several characteristics of this problem that make it uniquely challenging for automated detection. First, fake news is intentionally written to mislead readers, which makes it nontrivial to detect simply based on news content. The content of fake news is rather diverse in terms of topics, styles and media platforms, and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. For example, fake news may cite true evidence within the incorrect context to support a non-factual claim. Thus, existing hand-crafted and data-specific textual features are generally not sufficient for fake news detection. Other auxiliary information must also be applied to improve detection, such as knowledge base and user social engagements. Second, exploiting this auxiliary

information actually leads to another critical challenge: the quality of the data itself. Fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of corroborating evidence or claims. In addition, users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Effective methods to differentiate credible users, extract useful post features and exploit network interactions are an open area of research and need further investigations.

Methodology

Step 1: Feature Extraction

News content features describe the meta information related to a piece of news. A list of representative news content attributes are listed below:

- Source: Author or publisher of the news article
- Headline: Short title text that aims to catch the attention of readers and describes the main topic of the article
- Body Text: Main text that elaborates the details of the news story; there is usually a major claim that is specifically highlighted and that shapes the angle of the publisher
- Image/Video: Part of the body content of a news article that provides visual cues to frame the story.

Based on these raw content attributes, different kinds of feature representations can be built to extract discriminative characteristics of fake news. Typically, the news content we are looking at will mostly be linguistic-based and visual-based.

Step2: Model Construction

Since fake news attempts to spread false claims in news content, the most straightforward means of detecting it is to check the truthfulness of major claims in a news article to decide the news veracity. Knowledge-based approaches aim to use external sources to fact-check proposed claims in news content. The goal of fact-checking is to assign a truth value to a claim in a particular context. Fact-checking has attracted increasing attention, and many efforts have been made to develop a feasible automated fact-checking system. Existing fact-checking approaches can be categorized as expert-oriented, crowdsourcing-oriented, and computational-oriented.

Experimental Design

Datasets: Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs careful analysis of claims and additional evidence, context, and reports from authoritative sources. Generally, news data with annotations can be gathered in the following ways: Expert journalists, Fact-checking websites, Industry detectors, and Crowd-sourced workers

Evaluation Metrics: evaluate the performance of algorithms for fake news detection problem, various evaluation metrics have been used. In this subsection, we review the most widely used metrics for fake news detection. Most existing approaches consider the fake news problem as a classification problem that predicts whether a news article is fake or not:

- True Positive (TP): when predicted fake news pieces are actually annotated as fake news;
- True Negative (TN): when predicted true news pieces are actually annotated as true news;
- False Negative (FN): when predicted true news pieces are actually annotated as fake news;
- False Positive (FP): when predicted fake news pieces are actually annotated as true news.

By formulating this as a classification problem, we can define following metrics,

- $\text{Precision} = \frac{|T P|}{|T P| + |F P|}$
- $\text{Recall} = \frac{|T P|}{|T P| + |F N|}$
- $F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
- $\text{Accuracy} = \frac{|T P| + |T N|}{|T P| + |T N| + |F P| + |F N|}$

These metrics are commonly used in the machine learning community and enable us to evaluate the performance of a classifier from different perspectives. Specifically, accuracy measures the similarity between predicted fake news and real fake news.

C10. Feature Based Opinion Mining on Student Feedback

Problem Statement

Student feedback have the presence of huge amount of structured data like the grades, enrollment data, progression rates as well as unstructured data like student opinions expressed through surveys, web blogs, twitter, Facebook etc. It becomes highly time and resource consuming to summarize the information manually to reach data led conclusions and decisions. It is crucial to understand the patterns generated by the data like student feedback to effectively improve the performance of the institution and to create plans to enhance institutions' teaching and learning experience. Opinion Mining technique for classifying the students' feedback obtained during evaluation survey that is conducted every semester to know the feedback of students with respect to various features of teaching and learning such as module, teaching, assessments, etc. The extracted and preprocessed datasets can be subjected to various machine learning algorithm such as Support Vector Machine (SVM), Naïve Bayes (NB), K Nearest Neighbor (KNN) and Neural Networks (NN).

Background

Students feedback can help the lecturers understand their students learning behavior and improve teaching. Taking feedback can highlight different issues that the student may have with the lecture. One example of this is when the student does not understand part of the lecture or a specific example. Another example is when the lecturers' teaching pace is too fast or too slow. Feedback is usually collected at the end of the unit, but it is more beneficial taken in real-time. Collecting feedback has numerous benefits for the lecturer and their students, such as improvement in teaching and understanding student's learning behavior. Student's feedback improves communication between the lecturer and the students, allowing the lecturer to have an overall summary of the student's opinion. Student feedback can be collected using mobile phones and social media.

Feature based opinion mining deals with the extraction of the different features of the feedback of the student in an educational organization. It can be used to extract information from the student feedback about the teaching and learning methods adopted in an educational institute. Student feedback have the features like the grades, enrollment data, progression rates as well as unstructured data like student opinions expressed through surveys, web blogs, twitter, Facebook etc.

Student feedback is collected in form of responses to questions in a single sentence, it requires sentiment analysis in sentence level. In sentiment classification, machine learning methods have been used to classify each question as positive or negative. Testing of data is done based on training model which is classified using supervised learning algorithm. Evaluation of the total responses for every question and determine the polarity of feedback received in context of the question. The evaluation of response is purely data driven and hence simple while the classification of questions in form of natural language texts involves sentiment analysis. To test the model, collected data from students who posted their views in online discussion forums.

Methodology

Step 1: Data collection

This will involve collection of student feedback in the form of structured data like the grades, enrollment data, progression rates as well as unstructured data like student opinions expressed through surveys, web blogs, twitter, Facebook etc.

Step 2: Data Preprocessing

In this phase, the data is prepared for the analysis purpose which contains relevant information. Pre-processing and cleaning of data are one of the most important tasks that must be one before dataset can be used for machine learning. The real-world data is noisy, incomplete and inconsistent. So, it is required to be cleaned.

Step 3: Extraction of Feature Set/Training Data

In this phase, the cleaned data that is obtained from the data preprocessing phase is used to obtain the feature sets or training data of the student feedback. When we train to a classifier by taking maximum numbers of features, that contains all the irrelevant or redundant features can negatively affect the algorithm performance. So, it is required to carefully select the number and types of features that will be used to train the machine learning algorithms. Various feature selection techniques can be used for selecting features in the feature set/training data.

Feature set or training data can be prepared from the cleaned data by using any of the available techniques like bag of words, -gram, N-gram, POS, TOS tagging etc. The training data can also be prepared by providing them labels and then divide it into two classes like positive class and negative class. The feature sets and training

set that has obtained by using any of the above methods will be used for the implementation of machine learning algorithms.

Step 4: Implementation of Machine Learning Algorithm on Feature Set/Training Data

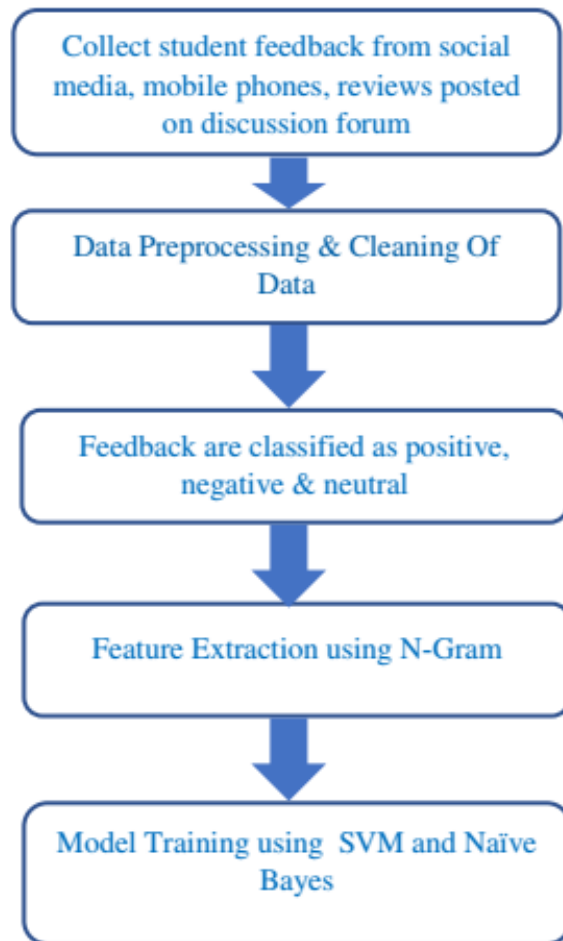
In the educational domain, Naïve Bayes and Support Vector Machines could be the best technique.

Naive classifier is a kind of probabilistic classifier. It is based on Bayes theorem with the assumptions that features between the feature sets are independent of each other. These classifiers are highly scalable. It is a simple method for developing a classifier. The model develops by Naïve Bayes classifier provide a class labels to the features of the feature set/training datasets.

SVM can be used for the analysis of data for both kind of supervised learning algorithms like classification and regression. When a set of training examples are given in which the sets are belongs to one or the other of two categories then SVM algorithm builds a model that will assigns unknown data the one or other categories' this way SVM is a non-probabilistic binary linear classifier. In SVM, representation of feature sets is done in the form of points in space, so that the features of the other categories are separated by clear space that is as wide as possible. Unknown feature is mapped and predicted to belongs to the space on which side of the gap it falls. It is shown as below

Step 5: Testing on Datasets

Testing of data is done based on training model which is classified using supervised learning algorithm. Evaluation of the total responses for every question and determine the polarity of feedback received in context of the question. The evaluation of response is purely data driven and hence simple while the classification of questions in form of natural language texts involves sentiment analysis. To test the model, collected data from students who posted their views in online discussion forums. Architecture is as follows:



Experimental Design

Dataset

Student feedback can be collected using mobile phones, social media and in form of responses to questions in a single sentence, from students who posted their views in online discussion forums.

Evaluation Measures

- **Accuracy:** Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made.

$$Accuracy = \frac{\text{Number of Correct predictions}}{\text{Total number of predictions made}}$$

- **Precision:** It is the number of correct positive results divided by the number of positive results predicted by the classifier.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

- **Recall:** It is the number of correct positive results divided by the number of *all* relevant samples (all samples that should have been identified as positive).

$$Precision = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Software and Hardware Requirements

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, jupyter notebook and libraries such as Tensorflow, and Keras will be utilized for this process.

References

- Khan, Khairullah, “Mining opinion components from unstructured reviews: A review,” Journal of King Saud University – Computer and Information Sciences, Vol. 26, 2014.
- Jyotsna Talreja Wassen, “Discovering Big Data Modelling for Educational World”, IETC Procedia - Social and Behavioral Sciences, pp:642 – 649, 2015
- Dirk T. Tempelaar, “In search for the most informative data for feedback generation: Learning analytics in a data-rich context”, :Journal of Computers in Human Behavior, Vol 47, pp: 157– 167, 2015.

- Beth Dietz-Uhler and Janet E. Hurn, “Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective”, Journal of Interactive Online Learning, Volume 12, Number 1, 2013.
 - Marie Bienkowski, Mingyu Feng, “Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics”, Department of Education, Office of Educational Technology: October 2012.
 - Pong Pien and Lillian Tsai, “Opinion mining and sentiment analysis : Foundations and trends in information retrieval”, Vol. 2, 2008.
7. Walaa Medhat, Ahmed Hassan, Hoda Korashy, “Sentiment analysis algorithms and applications: A survey”, Ain Shams Engineering Journal, Vol.5, 1093–1113, 2014.
 8. Kumar Ravi, Vadlamani Ravi, “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications”, Published in Knowledge-Based Systems Vol. 89 , 14–46, 2015.

C11. Malware Identification Using Deep Learning (NE)

Problem Statement

Deep learning (also known as **deep structured learning** or **hierarchical learning**) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. Deep learning models are loosely related to information processing and communication patterns in a biological nervous system, such as neural coding that attempts to define a relationship between various stimuli and associated neuronal responses in the brain. Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics and drug design where they have produced results comparable to and in some cases superior to human experts. In hierarchical Feature Learning, we extract will multiple layers of non-linear features and pass them to a classifier that combines all the features to make predictions. We are interested in stacking such very deep hierarchies of non-linear features because we cannot learn complex features from a few layers. It can be shown mathematically that for images the best features for a single layer are edges and blobs because they contain the most information that we can extract from a single non-linear transformation. To generate features that contain more information we cannot operate on the inputs directly, but we need to transform our first features (edges and blobs) again to get more complex features that contain more information to distinguish between classes.

Background

Most modern deep learning models are based on an artificial neural network, although they can also include propositional formulas or latent variables organized layer-wise in deep generative models such as the nodes in Deep Belief Networks and Deep Boltzmann Machines.

In deep learning, each level learns to transform its input data into a slightly more abstract and composite representation. In an image recognition application, the raw input may be a matrix of pixels; the first representational layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; and the fourth layer may recognize that the image contains a face. Importantly, a deep learning process can

learn which features to optimally place in which level *on its own*. (Of course, this does not completely obviate the need for hand-tuning; for example, varying numbers of layers and layer sizes can provide different degrees of abstraction.

The "deep" in "deep learning" refers to the number of layers through which the data is transformed. More precisely, deep learning systems have a substantial *credit assignment path* (CAP) depth. The CAP is the chain of transformations from input to output. CAPs describe potentially causal connections between input and output. For a feedforward neural network, the depth of the CAPs is that of the network and is the number of hidden layers plus one (as the output layer is also parameterized). For recurrent neural networks, in which a signal may propagate through a layer more than once, the CAP depth is potentially unlimited. No universally agreed upon threshold of depth divides shallow learning from deep learning, but most researchers agree that deep learning involves CAP depth > 2 . CAP of depth 2 has been shown to be a universal approximator in the sense that it can emulate any function. Beyond that more layers do not add to the function approximator ability of the network. The extra layers help in learning features.

Deep learning architectures are often constructed with a greedy layer-by-layer method. Deep learning helps to disentangle these abstractions and pick out which features improve performance.

Methodology

Step1: take some data

Step2: train a model on that data

Step3: use the trained model to make predictions on new data.

The process of training a model can be seen as a learning process where the model is exposed to new, unfamiliar data step by step. At each step, the model makes predictions and gets feedback about how accurate its generated predictions were. This feedback, which is provided in terms of an error according to some measure (for example distance from the correct solution), is used to correct the errors made in prediction.

Anomaly detection in network activities

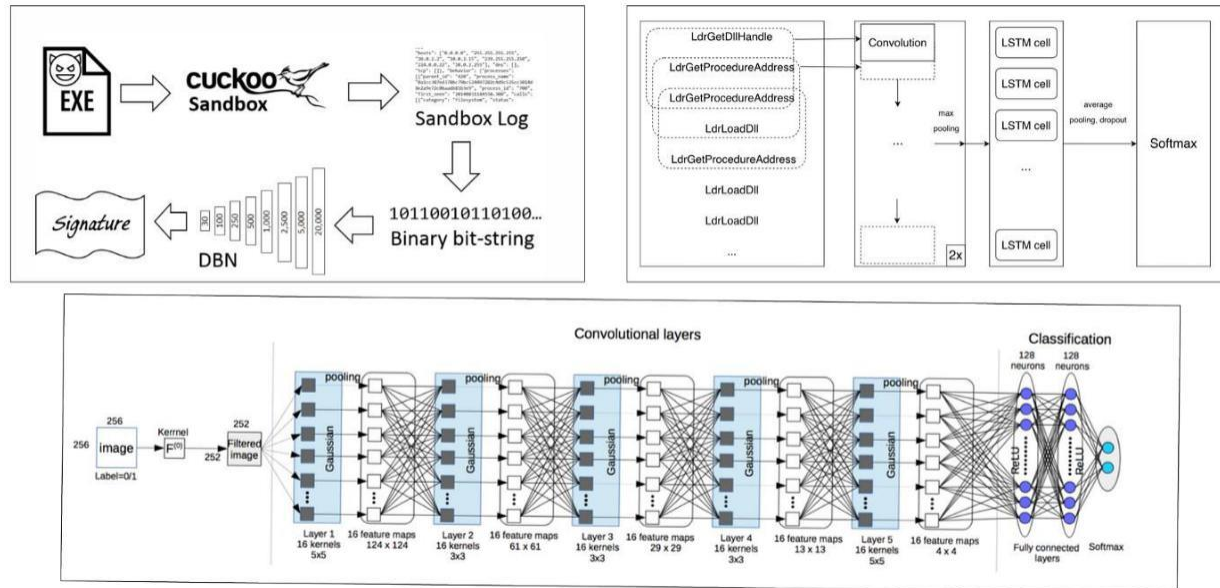


Fig 2: Architecture of Malware Detection using deep learning

Experimental Design

Dataset: All types of malwares.

Evaluation Measures: Measures such as accuracy and Mean Average Precision (MAP) will be computed by comparing the two different bounding boxes and ground truth boxes from the datasets.

Software and Hardware Requirements: Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as OpenCV, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the end-to-end version of CNN based object detection model.

C12. Spam SMS filtering Using Machine Learning (NE)

Problem Statement

Short Message Service (SMS) is one of the well-known communication services in which a message sends electronically. The lessening in the cost of SMS benefits by telecom organizations has prompted the expanded utilization of SMS. This ascent pulled in assailants, which have brought about SMS Spam problem. Spam messages include advertisements, free services, promotions, awards, etc. People are using the ubiquity of mobile phone devices is expanding day by day as they give a vast variety of services by reducing the cost of services. Short Message Service (SMS) is one of the broadly utilized communication service. In any case, this has prompted an expansion in mobile phones attacks like SMS Spam. In this problem, preliminary results are mentioned or explained herein based on Singapore based publically available datasets. This problem is further expanded using multiple background datasets.



Background

Many SMS Spam messages detection techniques are available these days to block spam messages and filtering spam messages. Few of which are mentioned below: - Gómez Hidalgo et al. assessed a few Bayesian based classifiers to identify mobile phone spam. In this problem, the researchers proposed the first two surely understood SMS spam datasets: the Spanish (199 spam and 1,157 ham) and English (82 spam and 1,119 ham) test databases. They have tried on them various messages portrayal techniques and machine learning calculations, as far as viability. The outcomes show that Bayesian separating methods can be successfully utilized to group SMS spam.

Hidalgo et al have analyzed that how Bayesian filtering technique can be used to detect SMS Spam. They have built two datasets one in English and another in Spanish. Their analysis shows that Bayesian filtering techniques that were earlier used in detecting email spam can also be used to block SMS Spam.

Methodology

Our framework is a combination of various techniques or methods like: - data collection, data pre-processing, and finally applying classifications methods on available SMS labeled dataset.

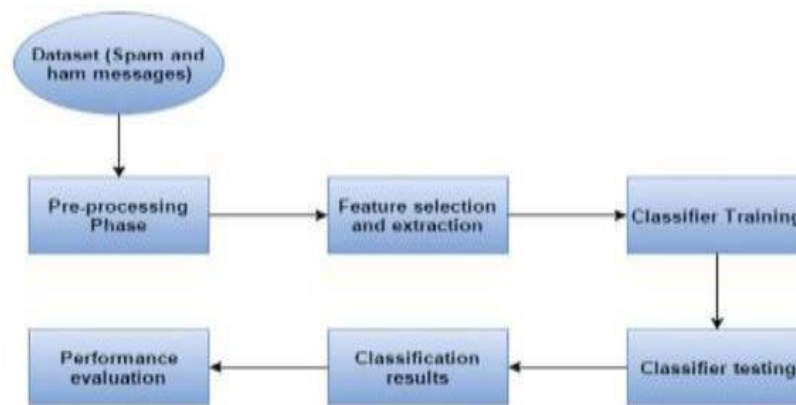


Fig1: System Architecture

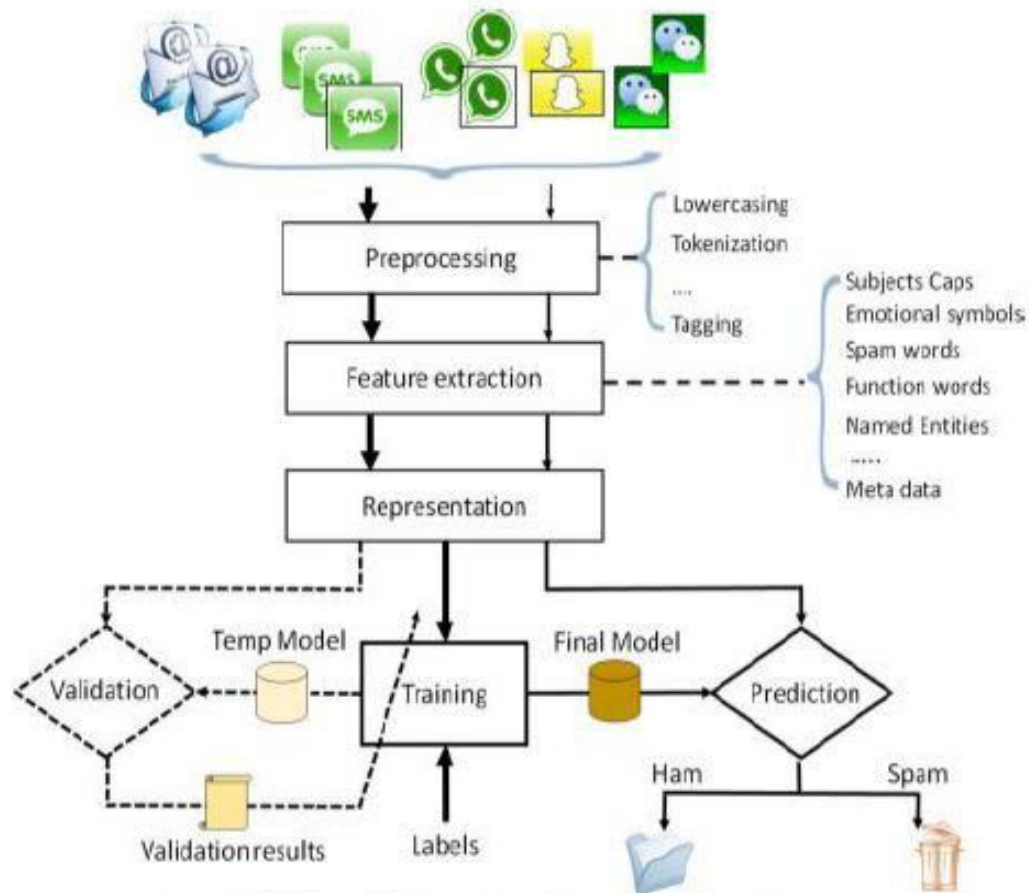


Fig 2: Flow Chart of SMS spam filtering [1]

Experimental Design

The SMS Spam Collection v.1 is a public set of SMS (text) labeled messages that have been collected for mobile phone spam research. It has one dataset composed by 5,574 English, real and non-encoded messages, tagged as legitimate (ham) or spam. The collection is free for all purposes, and it is publicly available at:

Links: 1. www.dt.fee.unicamp.br/~tiago/smsspamcollection/

- <https://archive.ics.uci.edu/ml/datasets/sms+spam+collection>

Name: - SMS Spam Collection v. 1 and SMS Spam Collection Data Set

When testing has been done by some researchers on the similar number of instances but written in Indian English our accuracy is degraded because of Indian writing style. Singapore English is different from Indian English because of different writing and speaking style. Our main objective following existing research is to develop such generalized model that could predict or filter the datasets on multiple backgrounds with a better accuracy level.

Our future goal should be test the results on large spam dataset on multiple backgrounds like (Singapore, American, Indian English etc.).

References

- Mobile Commons Blog.

<https://www.mobilecommons.com/blog/2016/01/how-textmessaging-will-change-for-the-better-in-2016/>

- http://www.academia.edu/2987380/SMS_Spam_Filtering_Methods_and_Data

C13 ->C15: Topic của TS. Hà Như Hằng

C13. Document Mapping Tool

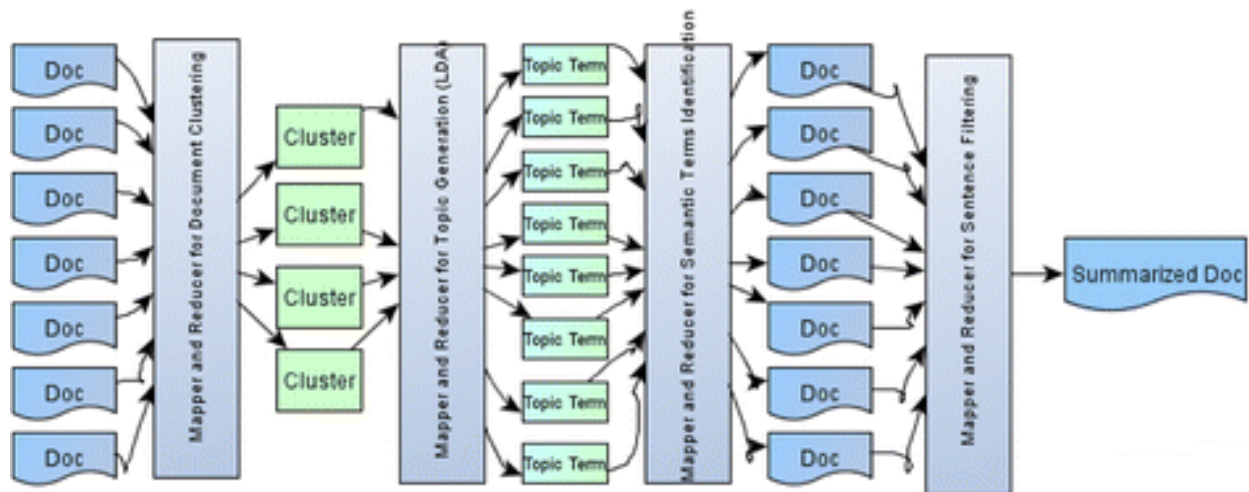
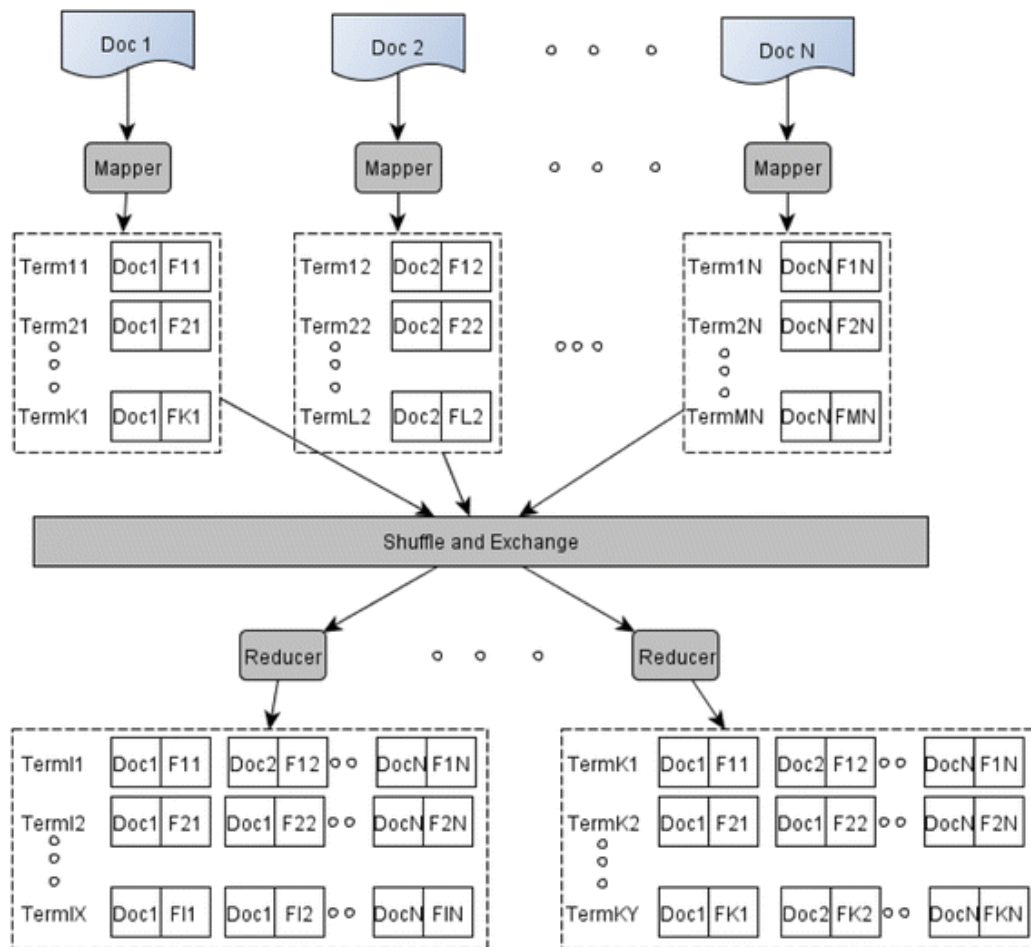
Hỗ trợ cho các nhà khoa học – viết các công trình nghiên cứu, các nhà quản lý – xử lý thông tin để ra quyết định.

Khi cần tìm kiếm thông tin, tài liệu liên quan đến chủ đề nào đó, thường mình sẽ download xuống là lưu nó vào trong 1 folder hoặc tách ra thành nhiều folder con theo kiểu như cây thư mục (Topic → subtopics)

Khi cần đọc thông tin phải mở từng file ra, như vậy cần mở nhiều files, rất bất tiện và không thể tìm được thông tin cần thiết 1 cách nhanh nhất, hoặc thay đổi các options khác nhau nếu thay đổi mục đích tìm kiếm.

Công cụ này sẽ hoạt động theo nguyên tắc như sau:

- Input: các văn bản có liên quan (Doc 1, 2, 3...)
- Process:
 - o Chọn option tìm kiếm: từ khóa
 - o Công cụ sẽ mining dữ liệu của từng văn bản (Doc 1, 2, 3...)
 - o Mỗi văn bản cho ra 1 bản đồ (Map) – trong đó sẽ chỉ ra mối liên kết với từ khóa
 - o Lọc các kết quả của từng văn bản – xây dựng 1 map chung
- Output: Clustering, Mapping ...



Có thể tham khảo thêm thông tin tại link sau:

<https://link.springer.com/article/10.1186/s40537-015-0020-5>

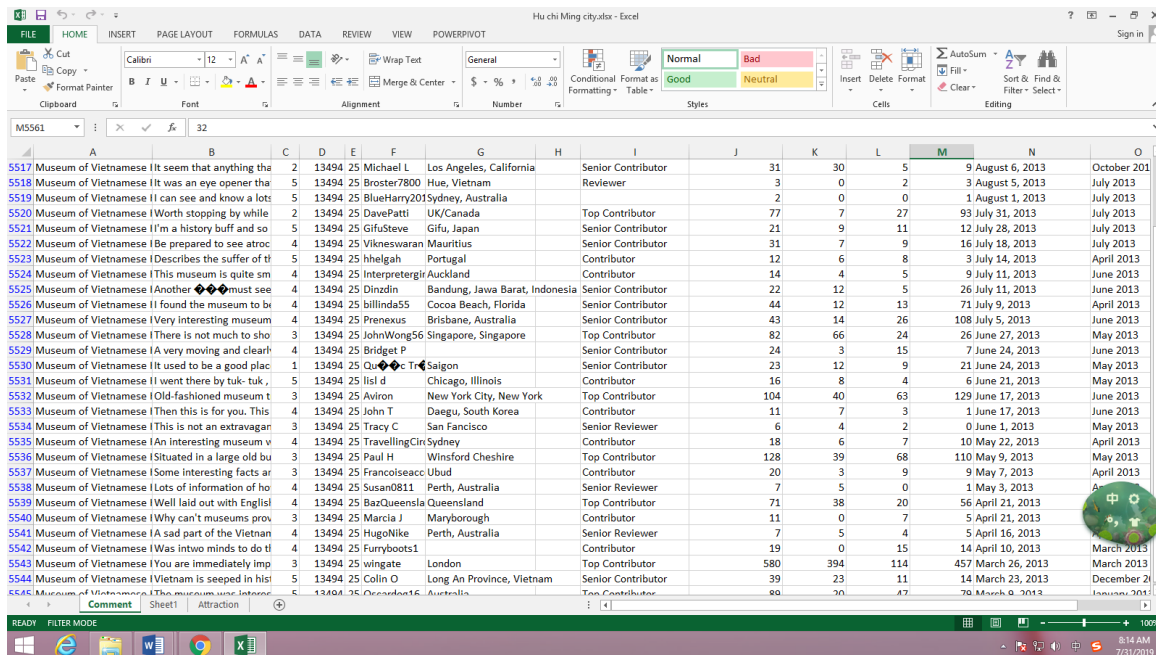
C14. Tool for content modelling

Hỗ trợ cho việc xử lý big data dạng text

Hiện tại có nhiều công cụ xử lý quantitative data (dữ liệu số) như Excel, SPSS, PLS, AMOS... Nhưng công cụ xử lý qualitative data (dữ liệu dạng chữ) còn hạn chế. Có xử lý nhưng chỉ đưa ra các dạng clusters – phân nhóm, phân loại, cần những công cụ đưa ra những mô hình tương quan giữa các nhóm được phân loại. Nhóm A, B ảnh hưởng như thế nào đến nhóm C (theo 1 số mô hình có sẵn, hay điều kiện ràng buộc).

Công cụ này yêu cầu:

- Input: có sẵn những file dữ liệu (vài chục ngàn cases) ví dụ như:

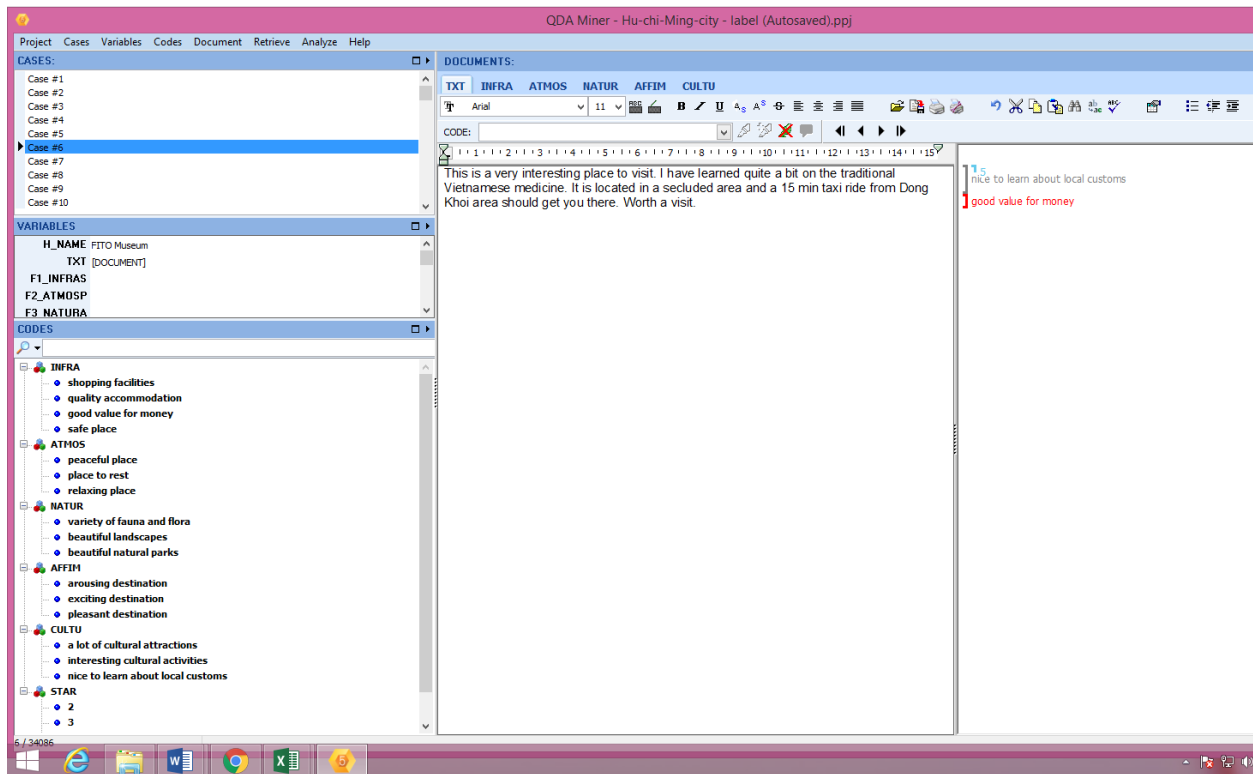


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5517	Museum of Vietnamese I	It seem that anything tha	2	13494	25	Michael L	Los Angeles, California		Senior Contributor		31	30	5	9 August 6, 2013	October 201
5518	Museum of Vietnamese I	It was an eye opener tha	5	13494	25	Broster7800	Hue, Vietnam		Reviewer		3	0	2	3 August 5, 2013	July 2013
5519	Museum of Vietnamese I	I can see and know a lots	5	13494	25	BlueHarry201	Sydney, Australia				2	0	0	1 August 1, 2013	July 2013
5520	Museum of Vietnamese I	Worth stopping by while	2	13494	25	DavePatti	UK/Canada		Top Contributor		77	7	27	93 July 31, 2013	July 2013
5521	Museum of Vietnamese I	I'm a history buff and so	5	13494	25	GifuSteve	Gifu, Japan		Senior Contributor		21	9	11	12 July 28, 2013	July 2013
5522	Museum of Vietnamese I	Be prepared to see atroc	4	13494	25	Vikneswaran	Mauritius		Senior Contributor		31	7	9	16 July 18, 2013	July 2013
5523	Museum of Vietnamese I	Describes the suffer of th	5	13494	25	hheglah	Portugal		Contributor		12	6	8	3 July 14, 2013	April 2013
5524	Museum of Vietnamese I	This museum is quite sm	4	13494	25	Interpretergii	Auckland		Contributor		14	4	5	9 July 11, 2013	June 2013
5525	Museum of Vietnamese I	Another must see	4	13494	25	Dinzlin	Bandung, Jawa Barat, Indonesia		Senior Contributor		22	12	5	26 July 11, 2013	June 2013
5526	Museum of Vietnamese I	I found the museum to bi	4	13494	25	billinda55	Cocoa Beach, Florida		Senior Contributor		44	12	13	71 July 9, 2013	April 2013
5527	Museum of Vietnamese I	Very interesting museum	4	13494	25	Prexenus	Brisbane, Australia		Senior Contributor		43	14	26	108 July 5, 2013	June 2013
5528	Museum of Vietnamese I	There is not much to sho	3	13494	25	JohnWong56	Singapore, Singapore		Top Contributor		82	66	24	26 June 27, 2013	May 2013
5529	Museum of Vietnamese I	A very moving and clear	4	13494	25	Bridget P			Senior Contributor		24	3	15	7 June 24, 2013	June 2013
5530	Museum of Vietnamese I	It used to be a good plac	1	13494	25	Quoc Tr	Saigon		Senior Contributor		23	12	9	21 June 24, 2013	May 2013
5531	Museum of Vietnamese I	I went there by tuk-tuk	5	13494	25	lisl d	Chicago, Illinois		Contributor		16	8	4	6 June 21, 2013	May 2013
5532	Museum of Vietnamese I	Old-fashioned museum t	3	13494	25	Aviron	New York City, New York		Top Contributor		104	40	63	129 June 17, 2013	June 2013
5533	Museum of Vietnamese I	Then this is for you. Thi	4	13494	25	John T	Daegu, South Korea		Contributor		11	7	3	1 June 17, 2013	June 2013
5534	Museum of Vietnamese I	This is not an extravagan	3	13494	25	Tracy C	San Francisco		Senior Reviewer		6	4	2	0 June 1, 2013	May 2013
5535	Museum of Vietnamese I	An interesting museum v	4	13494	25	TravellingCir	Sydney		Contributor		18	6	7	10 May 22, 2013	April 2013
5536	Museum of Vietnamese I	Situated in a large old bu	3	13494	25	Paul H	Winsford Cheshire		Top Contributor		128	39	68	110 May 9, 2013	May 2013
5537	Museum of Vietnamese I	Some interesting facts ar	3	13494	25	Francoiseacc	Ubud		Contributor		20	3	9	9 May 7, 2013	April 2013
5538	Museum of Vietnamese I	Lots of information of ho	4	13494	25	Susan0811	Perth, Australia		Senior Reviewer		7	5	0	1 May 3, 2013	April 2013
5539	Museum of Vietnamese I	Well laid out with Englis	4	13494	25	BazQueensla	Queensland		Top Contributor		71	38	20	56 April 21, 2013	April 2013
5540	Museum of Vietnamese I	Why can't museums prov	3	13494	25	Marcia J	Maryborough		Contributor		11	0	7	5 April 21, 2013	April 2013
5541	Museum of Vietnamese I	A sad part of the Vietna	4	13494	25	HugoNike	Perth, Australia		Senior Reviewer		7	5	4	5 April 16, 2013	March 2013
5542	Museum of Vietnamese I	You into minds to do th	4	13494	25	Funnyboots1			Contributor		19	0	15	14 April 10, 2013	March 2013
5543	Museum of Vietnamese I	You are immediately imp	3	13494	25	wingate	London		Top Contributor		580	394	114	457 March 26, 2013	March 2013
5544	Museum of Vietnamese I	Vietnam is seeped in his	5	13494	25	Colin O	Long An Province, Vietnam		Senior Contributor		39	23	11	14 March 23, 2013	December 2
5545	Museum of Vietnamese I	This museum was interest	6	13494	25	Oceandad16	Australia		Top Contributor		90	20	47	70 March 9, 2013	January 2013

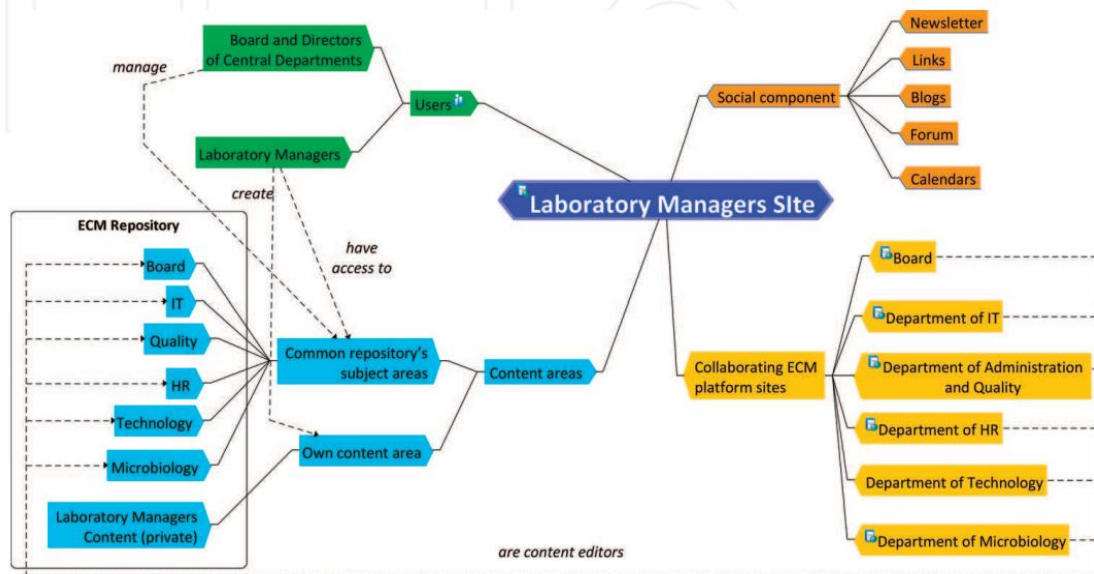
people_wiki.csv - Excel

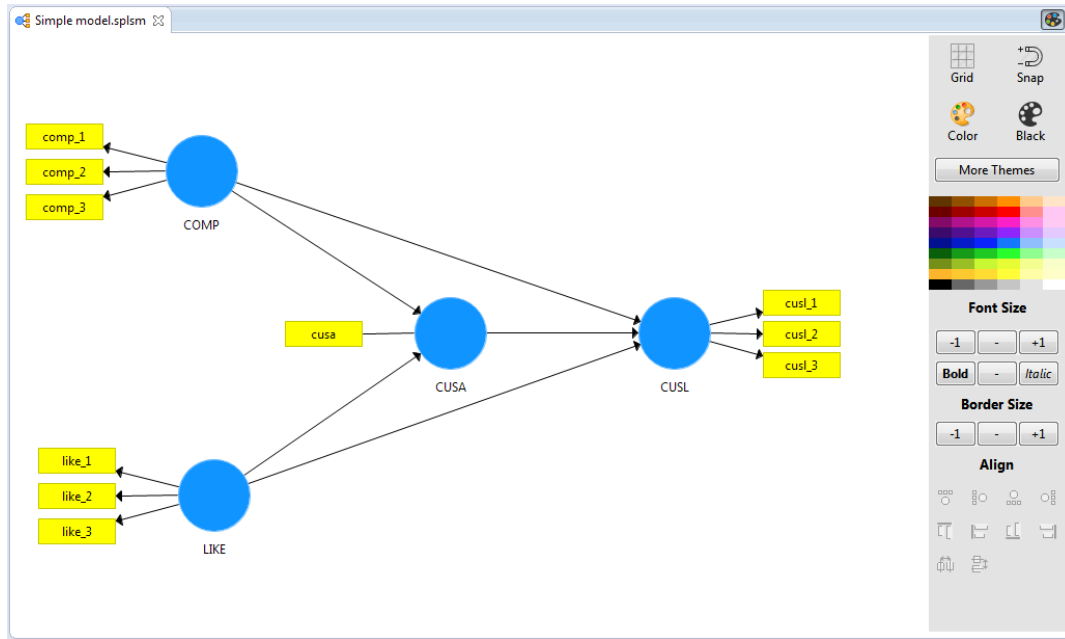
URI	name	text
<http://dl.Digby Mordigby morrell born 10 october 1979 is a former australian rules footballer who played with the kangaroos and carlton in the australian football league affrom western australia morrell played his early senior football for west perth hi		
<http://dl.Alfred J. L. alfred j lewy aka sandy lewy graduated from university of chicago in 1973 after studying psychiatry pharmacology and ophthalmology he is a full professor and vicechair of the department of psychiatry at ohsu oregon health science u		
<http://dl.Harpdog E harpdog brown is a singer and harmonica player who has been active in canadas blues scene since 1982 hailing from vancouver he crossed tens of thousands of miles playing club dates and festivals in canada the northwestern united		
<http://dl.Franz Rottfranz rotti nsteiner born in waidmannsfeld lower austria austria on 18 january 1942 is an austrian publisher and critic in the fields of science fiction and the fantasticrottensteiner studied journalism english and		
<http://dl.G-Enka henry krivits born 30 december 1974 in tallinn better known by his stagename genka is an estonian rapper and record producergenka started rapping in 1996 along with revo and dj paul oja who was genkas schoolmate together they s		
<http://dl.Sam Hend sam henderson born october 18 1969 is an american cartoonist writer and expert on american comedy historyhenderson was born in woodstock new york he attended boiceville new yorks oteora high school graduating in 1987 and t		
<http://dl.Aaron LaAaron lacrate is an american music producer recording artist dj fashion designer of milkcrate records and milkcrate clothing and a film director he was born in baltimore maryland and grew up in highlandtown east baltimoreaaron lac		
<http://dl.Trevor Feitrevor ferguson aka john farrow born 11 november 1947 is a canadian novelist who lives in hudson quebec he is the author of nine novels and four plays he has been called canadas best novelist both in books in canada and the toron		
<http://dl.Grant Neil grant nelson born 27 april 1971 in london also known as wishdokka bump flex and nng is an english dj remixer and record producernelson is heralded as one of the godfathers of uk garage due to his numerous club hits on his nice n ri		
<http://dl.Cathy Can cathy caruth born 1955 is frank h t rhodes professor of humane letters at cornell university and is appointed in the departments of english and comparative literature she taught previously at yale and at emory university where she h		
<http://dl.Sophie Crisophia violet sophie crumb born september 27 1981 is an americanfrench comics artist she is the daughter of underground comic artists robert crumb and aline kominskycrumb was born in woodland california and lived in the r		
<http://dl.Jenn Ashujenn ashworth is an english writer she was born in 1982 in preston lancashire she has graduated from cambridge university and the manchester centre for new writing she previously worked as a librarian in a mens prisonshe founded		
<http://dl.Jonathan Jonathon hoefler born august 22 1970 is an american typeface designer hoefler pronounced heffler advanced the hoefler type foundry in 1989 a type foundry in new york in 1999 hoefler began working with type designer tobias freerej		
<http://dl.Anthony anthony fitzhardingue gueterbock 18th baron berkeley and baron gueterbock obe ceng miche hon fimeche hon disc frsa fdt born 20 september 1939 also known as tony berkeley is a british labour politician he is both an english heredita		
<http://dl.David Che david chemushenko born june 1963 in calgary alberta is a politician professional speaker sustainability consultant and documentary filmmaker in ontario canada he was elected to ottawa city council in the 2010 municipal election an		
<http://dl.Joerg Stei joerg steineck is a german filmmaker editor and graphic designerartist who is based in berlinduring his study of film and design he produced a few short documentaries and fictional films in his films he often explores subcultures sut		
<http://dl.Andrew P fr andrew pinsent born 19 august 1966 is research director of the ian ramsey centre for science and religion at oxford university a member of the theology faculty a research fellow of harris manchester college and a catholic priest of		
<http://dl.Paddy Dui paddy dunne was a gaelic football player from park in county laoishe played for many years on the laois senior football team in the centre half back position and was widely regarded as one of the outstanding players in ireland of the		
<http://dl.Alexandri alexandros mouzas born 1962 is a greek composer he studied composition with theodore antoniou advanced theory with haris xantheadakis and electronic music with dimitris kamarotos he has composed for various ensembles and		
<http://dl.John Angu john angus campbell born march 10 1942 in portland oregon usa is a retired american professor of rhetoric and is a fellow of the center for science and culture a branch of the discovery institute a conservative christian think tank and		
<http://dl.Chris Bats chris batstone was the 20002002 lead singer of the thirdwave ska band suburban legends he joined the band in 2000 as the replacement for tim maurer who had left the band after the recording of origin edition chris first and only off		
<http://dl.Ceiron The ceiron thomas born 23 october 1983 is a welsh rugby union footballer currently playing for st ives rfc after signing from the cornish pirates in june 2012 his previous clubs are leeds carnegie and the scarlets his regular position is fly		
<http://dl.Adel Selli adel sellimi arabic was born on 16 november 1972 in the bab jedid district of tunis as a child he drew inspiration from the 1978 world cup team who became the first african nation to win a world cup match at the age of		
<http://dl.Faith Solo faith soloway born march 28 1964 is an american folkrock musician comedic performer and creator of several folkrock musicals based in the boston area she studied theater at indiana university and performed improv		
<http://dl.Tom Jenni tom jennings born 1955 as thomas daniel jennings in boston massachusetts is a los angelesbased artist and technician he is the creator of fidonet the first message and file networking system for bbcs originally the fidonet		
<http://dl.Vic Stasiul victor john stasiuk born may 23 1929 is a retired canadian professional ice hockey left winger and a former nhl head coach stasiuk played junior hockey in his native lethbridge alberta before signing with the chicago black hawk		
<http://dl.Anthony anthony caruana born 2 january 1968 is a melbourne australia based technology journalist who writes for australian macworld the australian broadcasting commission several titles from next media it wire and many other publication		
<http://dl.Ian Mitchi ian mitchell is a scottish author who grew up mainly in south africa he is the author of isles of the west a hebridean voyage and isles of the north a voyage to the realms of the norse both books are concerned with apparentlyvirtuous		
<http://dl.Leon Hapi leon duane haggood born 7 august 1979is an english amateur footballer he has represented england schoolboyshaggood was born in torquay and began his career as an apprentice with torquay united scoring on his league debut in s		

- Process:
 - o Đưa dữ liệu vào công cụ: mỗi dữ liệu có 1 mục đích phân tích riêng, các mục đích này liên quan đến 1 số biến nhất định (variables/factors)
 - o Đặt tên cho từng cụm từ khóa trong từng dòng dữ liệu theo variable cho sẵn:
 - Ví dụ: Variable tính lỗi cuộn --- Code 1: ATT1; đọc từng dòng dữ liệu, nếu xuất hiện từ liên quan đến lỗi cuộn thì đẩy vào nhóm Code 1



- Sau đó Analyze
- Output: Mô hình, có thể dạng Mindmap, hoặc quan hệ các variable





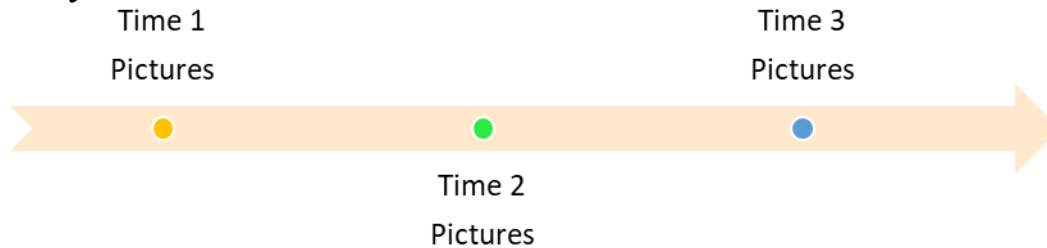
C15. Moments of life (App)

Hỗ trợ phân loại ảnh trong folder ảnh trên điện thoại

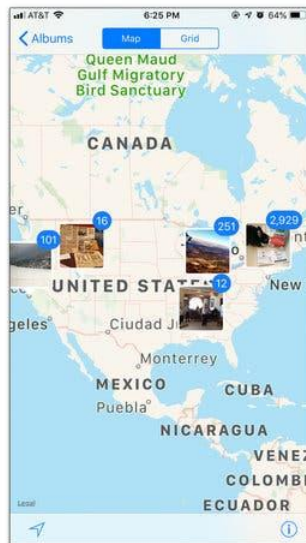
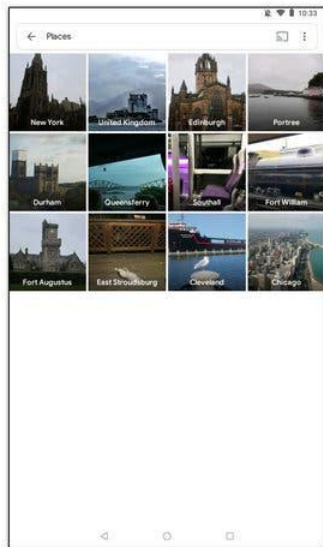
Trong 1 bức ảnh thường có các yếu tố: Thời gian, địa điểm, con người

Dựa vào thông tin bức ảnh chụp có trong folder có thể tạo thành ra những sản phẩm theo kiểu:

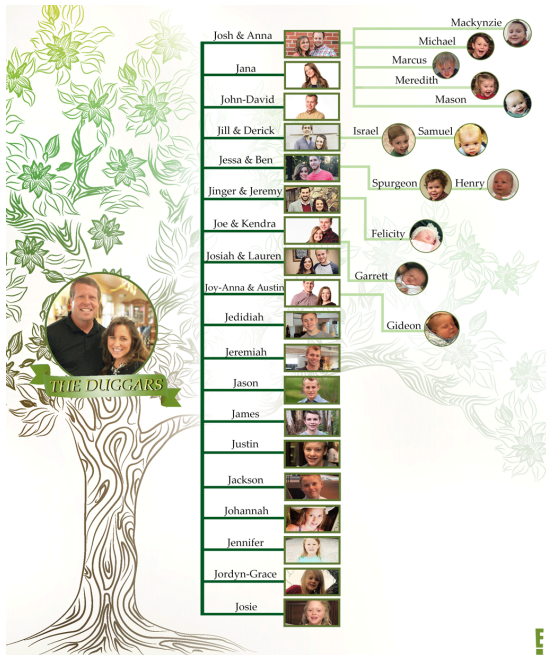
- *Story:*



- *Location:*



- *Relationship:*



- *Loại khác...*

Topic	C16. Nghiên cứu Deep learning & xây dựng model phân biệt các lớp động vật khác nhau
Mã đề tài	C16
Công cụ, công nghệ sử dụng	Python, R
Mục tiêu nghiên cứu và kết quả cần đạt được	Nghiên cứu các hidden layer, mô hình nào là hợp lý nhất. Tạo ra nhiều mô hình và chọn mô hình có độ chính xác cao nhất để áp dụng cho từng lớp động vật cụ thể.
Mô tả nội dung yêu cầu	<p>///Mô tả rõ bài toán, chủ đề mà sinh viên cần biết để làm</p> <p>(1) Đọc các tài liệu tham khảo để hiểu vấn đề cần thực hiện; Nghiên cứu và trình bày deep learning, các thách thức</p> <p>(2) Trình bày cách tiếp cận Mạng neural: Phương pháp tiếp cận, xây dựng mô hình, sinh dữ liệu tổ hợp tự động</p> <p>(3) Phát triển tool để tạo ra các model mạng neural khác nhau; Đánh giá kết quả về các model, công sức xây dựng test case và test data</p> <p>(4) Xây dựng phương án / ý tưởng cải tiến dựa trên các phân tích sâu hơn về thuật toán</p> <p>(5) Phát triển/hiệu chỉnh mã nguồn chương trình áp dụng các kỹ thuật lập trình đã được học</p> <p>(6) Tạo tài liệu về phần mềm theo đúng dạng chuẩn</p>
Yêu cầu khác	Lấy từ nguồn Git-Hub và bàn bạc với giảng viên hướng dẫn để có thêm các dữ liệu thử nghiệm bổ sung.
Các tài liệu và thư viện chi tiết kèm theo để giao cho sinh viên/ Tài liệu tham khảo, tutorial	<p>[1] https://www.kaggle.com/c/dogs-vs-cats chứa các data về lớp chó và mèo</p> <p>[2]https://towardsdatascience.com/cat-or-dog-image-classification-with-convolutional-neural-network-d421a9363c7a code python xây dựng model</p>

Topic	Nghiên cứu và đề xuất giải pháp mô hình hóa việc tích hợp dữ liệu
Mã đề tài	C17
Công cụ, công nghệ sử dụng	Java/XML/UML/JavaScripts
Mục tiêu nghiên cứu và kết quả cần đạt được	Nghiên cứu và đưa ra giải pháp về việc tích hợp dữ liệu từ nhiều nguồn dữ liệu khác nhau để thu thập vào kho dữ liệu, hỗ trợ xây dựng hệ hỗ trợ ra quyết định.
Mô tả nội dung yêu cầu	<p>(1) . Nghiên cứu các giải pháp tích hợp dữ liệu đã có</p> <p>(2). Nghiên cứu các về phương pháp về mô hình hóa dữ liệu</p> <p>(3). Nghiên cứu các công nghệ, công cụ và các ngôn ngữ: Java/XML/UML/JavaScripts...</p> <p>(4). Nghiên cứu về mô hình dữ liệu đa chiều (multidimensional data model)</p> <p>(5). Giải pháp chuyển đổi mô hình</p>
Yêu cầu khác	Trao đổi giảng viên hướng dẫn để có thêm tài liệu và định hướng nghiên cứu
Các tài liệu và thư viện chi tiết kèm theo để giao cho sinh viên/ Tài liệu tham khảo, tutorial	Mentor sẽ gửi cho sinh viên
Ghi chú	Đề tài này theo hướng nghiên cứu khoa học, dự kiến phát triển thành bài báo hoặc đăng ký công trình nghiên cứu khoa học cấp trường.

Topic	Xây dựng ứng dụng chatbot hỗ trợ tư vấn bán quần áo
Mã đề tài	C18
Công cụ, công nghệ sử dụng	Java/Json/thư viện xử lý ngôn ngữ tự nhiên Nghiên cứu thêm về Machine learning để vận dụng trong quá trình xây dựng chatbot
Mục tiêu nghiên cứu và kết quả cần đạt được	Xây dựng ứng dụng Chatbot cho các website của các cửa hàng bán áo quần. Khách hàng chỉ cần truy cập vào các trang web của cửa hàng là có thể chọn được áo/quần mong muốn.
Mô tả nội dung yêu cầu	(1) . Nghiên cứu các công nghệ: java/json... (2). Nghiên cứu các về Machine learning (3). Sử dụng các công cụ/mã nguồn đã có để áp dụng trong xây dựng chatbot.
Yêu cầu khác	Trao đổi giảng viên hướng dẫn để có thêm tài liệu và định hướng nghiên cứu
Các tài liệu và thư viện chi tiết kèm theo để giao cho sinh viên/ Tài liệu tham khảo, tutorial	Mentor sẽ gửi cho sinh viên

Topic	Hệ thống nhân viên phục vụ ảo
Mã đề tài	C19
Công cụ, công nghệ sử dụng	
Mục tiêu nghiên cứu và kết quả cần đạt được	Xây dựng 1 hệ thống nhân viên phục vụ ảo tại chuỗi các quán café hay trà sữa tích hợp với hệ thống máy POS của quán để marketing khách hàng tốt nhất
Mô tả nội dung yêu cầu	<ul style="list-style-type: none"> - Tại quầy order sẽ có 1 hệ thống camera nhận diện và thu thập thông tin khách hàng. Hệ thống sẽ chào hỏi khách và xác định hành vi (thói quen) gọi thức uống của khách, để từ đó hệ thống đưa ra những gợi ý phù hợp với nhu cầu của khách - Hệ thống hỗ trợ giới thiệu và gợi ý món phục vụ khách một cách tốt nhất, tạo sự quan tâm, thân thiện của nhà hàng đối với khách hàng

