**ABSTRACT:**

**My project is about Expert System Classify and identify documents**

## One of the widely used natural language processing task in different business problems is “To Identify and Classify Document”. The goal of this project is to identify and classify the documents into one or more defined categories. In this we identify documents based on the file name for example, if the client don’t know the type of document he/she have and enter the file name as “xyz.doc”, then we are going to help them by saying the type of file it is for the above example the document type is a MS Word document. Same goes with classification of documents if the client enters the list of documents and he wants to segregate particular type of documents then he enters the type of documents then we are going to display list of documents of that type.

# ADVANTAGES OF This PROJECT:

# 1.We can easily identify the document is of what type.

# 2.And it will be easy to search

# 3.Time complexity will be less

# 4.We can directly interact with the system instead of searching

# 5.We can separate files easily based on their extensions

# 6.Human Errors will be reduced by using this expert system

# 7. This provide strategic and comparative advantages

**INTRODUCTION:**

Document Classification is an example of supervised machine learning task since a labelled dataset containing Document documents and their labels is used for train a classifier. An end-to-end Document Classification pipeline is composed of three main components:

**1.Dataset Preparation:**The first step is the Dataset Preparation step which includes the process of loading a dataset and performing basic pre-processing. The dataset is then splitted into train and validation sets.

**2. Feature Engineering:**The next step is the Feature Engineering in which the raw dataset is transformed into flat features which can be used in a machine learning model. This step also includes the process of creating new features from the existing data.

**3. Model Training:**The final step is the Model Building step in which a machine learning model is trained on a labelled dataset.

**4. Improve Performance of Document Classifier:**In this article, we will also look at the different ways to improve the performance of Document classifiers.

**Expert System:**

**Expert system** is a computer system that emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as if–then rules rather than through conventional procedural code. The first expert systems were created in the 1970s and then proliferated in the 1980s. Expert systems were among the first truly successful forms of artificial intelligence(AI) software.

An expert system is divided into two subsystems: the inference engine and the knowledge base. The knowledge base represents facts and rules. The inference engine applies the rules to the known facts to deduce new facts. Inference engines can also include explanation and debugging abilities.

## Capabilities of Expert Systems:

The expert systems are capable of −

* Advising
* Instructing and assisting human in decision making
* Demonstrating
* Deriving a solution
* Diagnosing
* Explaining
* Interpreting input
* Predicting results
* Justifying the conclusion
* Suggesting alternative options to a problem

### Advantages:

Expert systems use information technology to gain and use human expertise. Obviously, this can be very beneficial to organizations. Expert Systems can:

1. Provide answers for decisions, processes and tasks that are repetitive
2. Hold huge amounts of information
3. Minimize employee training costs
4. Centralize the decision making process
5. Make things more efficient by reducing the time needed to solve problems
6. Combine various human expert intelligences
7. Reduce the number of human errors
8. Provide strategic and comparative advantages that may create problems for competitors
9. Look over transactions that human experts may not think of

**DISDVANTAGES:**

However, there are also disadvantages to expert systems, such as:

1. No common sense used in making decisions
2. Lack of creative responses that human experts are capable of
3. Not capable of explaining the logic and reasoning behind a decision
4. It is not easy to automate complex processes
5. There is no flexibility and ability to adapt to changing environments
6. Not able to recognize when there is no answer

## Components of Expert Systems:

The components of ES include −

* Knowledge Base
* Inference Engine
* User Interface

**Source code:**

import random

print('whats your name ')

a=str(input())

print('\n')

print("hello "+a+", Welcome to expert system of documents")

print('\n')

print('what you want me to do')

ques=['identify documents','Identify documents','classify documents','Classify Documents']

s=str(input())

print('\n')

type=0

if ques[0] in s or ques[1] in s:

type=1

elif ques[2] in s or ques[3] in s:

type=2

if(type==1):

print('enter the file name along with extension: ')

fn=str(input())

l=len(fn)

for i in range(0,l):

if(fn[i]=='.'):

j=i

if(fn[(j+1):]=='exe'):

print('your file is a extension file')

elif(fn[(j+1):]=='doc' or fn[(j+1):]=='docx'):

print('your file is a MS Word file')

elif(fn[(j+1):]=='html'):

print('your file is a html file')

elif(fn[(j+1):]=='ppt' or fn[(j+1):]=='pptx'):

print('your file is a MS Powerpoint file')

elif(fn[(j+1):]=='py'):

print('your file is a python file')

elif(fn[(j+1):]=='java' or fn[(j+1):]=='class'):

print('your file is a java file')

elif(fn[(j+1):]=='c'):

print('your file is a c program file')

elif(fn[(j+1):]=='pdf'):

print('your file is a pdf file')

elif(fn[(j+1):]=='xls'):

print('your file is a MS Excel file')

elif(fn[(j+1):]=='zip'):

print('your file is a PC Zip Compressed Archive')

else:

print('enter correct file name along with extension')

if(type==2):

str = input("enter the file names seperated by comma : ")

arr = str.split(",")

exe=[]

doc=[]

html=[]

ppt=[]

py=[]

java=[]

c=[]

pdf=[]

xcel=[]

zip1=[]

for name in arr:

le=len(name)

for i in range(0,le):

if(name[i]=='.'):

j=i

if(name[(j+1):]=='exe'):

exe.append(name)

elif(name[(j+1):]=='doc' or name=='docx'):

doc.append(name)

elif(name[(j+1):]=='html'):

html.append(name)

elif(name[(j+1):]=='ppt' or name[(j+1):]=='pptx'):

ppt.append(name)

elif(name[(j+1):]=='py'):

py.append(name)

elif(name[(j+1):]=='java' or name[(j+1):]=='class'):

java.append(name)

elif(name[(j+1):]=='c'):

c.append(name)

elif(name[(j+1):]=='pdf'):

pdf.append(name)

elif(name[(j+1):]=='xls'):

xcel.append(name)

elif(name[(j+1):]=='zip'):

zip1.append(name)

else:

print('enter correct file with comma')

print('1.Executabale')

print('2.Ms word')

print('3.html fle')

print('4.powepoint')

print('5.python file')

print('6.java file')

print('7.c file')

print('8.pdffile')

print('9.excel file')

print('10.Zip compressed Archive')

i=int(input('enter what type of file u want to classify (i.e., between 1 to 10): '))

if(i==1):

print("extension documents are",exe)

elif(i==2):

print("MS word documents are",doc)

elif(i==3):

print("html documents are",html)

elif(i==4):

print("MS Powerpoints are",ppt)

elif(i==5):

print("Python files are",py)

elif(i==6):

print("java documents are",java)

elif(i==7):

print("c files are",c)

elif(i==8):

print("pdf documents are",pdf)

elif(i==9):

print("MS Excel documents are",xcel)

elif(i==10):

print("Zip files are",zip1)

**Output:**

**Output for identifying of documents:**

A screenshot of a computer program

Description automatically generated

**Output for classifying of documents:**

**A screenshot of a computer program

Description automatically generated**

**CONCLUSION:**

Here we declare that anyone can successfully identify and classify documents through this project by using expert system.so, the work will become easy for the clients who wants to identify or classify the documents.