

eDrug

by Team 4 (StrikerZ):

Venkata Vamsi Krishna Bhuvanam

Priyadarsini Nidadavolu

Bhulakshmi Makkena

Tej Kumar Yentrapragada

TABLE OF CONTENTS

1	Motivation.....	3
2	Objectives.....	4
3	Expected Outcome.....	5
4	Domain Chosen.....	6
5	Data Collection.....	7
6	Tasks and Features.....	8
7	Implementation	9
7.1	Architecture & UML Diagrams:.....	9
7.1.1	Sequence Diagram:	10
7.1.2	Class Diagram:.....	11
7.2	Workflow:	12
7.3	Existing Services:.....	13
7.3.1	First Increment:.....	13
7.3.2	Second Increment:.....	16
7.4	Feature Implemented:.....	29
8	Project Management	34
8.1	Contribution:.....	34
8.2	ZenHub and GitHub:	35
8.2.1	First Increment:.....	35
8.2.2	SECOND INCREMENT:.....	37
8.3	Concerns/Issues:	38
8.4	Future Work:.....	39

1 MOTIVATION

The motivation of building a recommendation system for personalized treatment is to diagnosis a disease based on symptoms provided by the users and then recommend a suitable and appropriate drug for the user based on few criteria's like past behavior and historical data of past similar symptoms of the patients along with the drug allergies that the patient has in the past.

Our system will recommend few drugs to select from a list which are approved by the pharmacist. This option will help a user to stream into the list of interest across more than one domain. The system would be able to return a suggestive drug in the list using various technologies that we plan on implementing.

2 OBJECTIVES

The main objective of the recommendation system for E-drug is to recommend a drug to the user based on the symptoms that he provided.

Our system takes user symptoms as input and it analyses the input with the types of diseases and produces a dataset with various possible diseases.

It analysis the type of diseases and ultimately recommends an appropriate drug based on the following criteria:

- Past Behavior of the patient
- Historical data of the past similar symptoms of the patients
- Drug allergies
- Drug similarities that the doctor recommended previously

Our recommendation system would return the dataset of list of recommended drugs with a hyperlink which leads to the composition, availability of the drug in various stores, the cost etc. This would help the user to view the list and can go and check for nearby pharma and buy it.

Machine learning and ALS algorithms run in the background and the result would be an recommended drug dataset.

3 EXPECTED OUTCOME

At the end, this machine should be able to take the dataset and recommend the patient/doctor with the suitable drug without any future side effects.

4 DOMAIN CHOSEN

For this project, Recommendation system will help a lot to analyze the data of symptoms for diseases and its drug compositions.

Also we have used a Spark Core Libraries and Spark Machine Learning Libraries.

5 DATA COLLECTION

Static Data:

<http://human-phenotype-ontology.github.io/downloads.html>

<https://www.nlm.nih.gov/medlineplus/encyclopedia.html>

Hospital Charge Data:

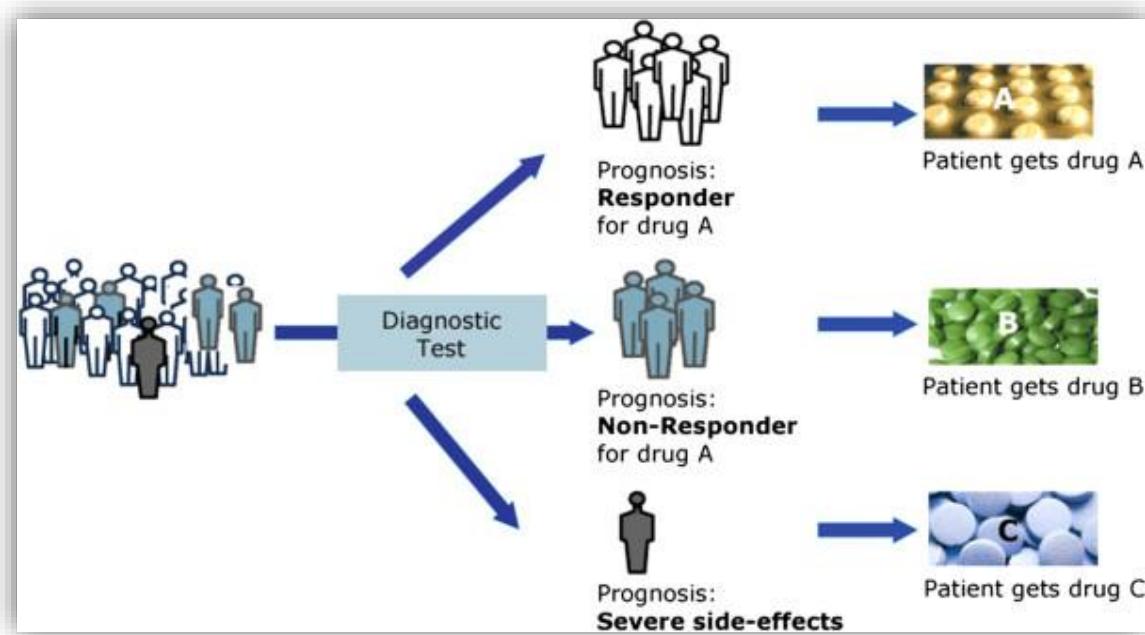
<https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>

6 TASKS AND FEATURES

This project definitely need a lot of data and to analyze that and get to a strong knowledge base. Using that knowledge, the machine will recommend the drug. Here there will be a lot of features to be considered. For example, first the machine will analyze the symptoms and different types of diseases. This machine will take a lots of criteria into consideration like past behavior of the patient, historical data of other patients and their acceptance or rejection of the drug. Also this machine will analyze the drug allergies of a particular patient and based on the finalized knowledge created, the machine will recommend some drugs, so that the doctors can have a look and prescribe particular drugs to patients.

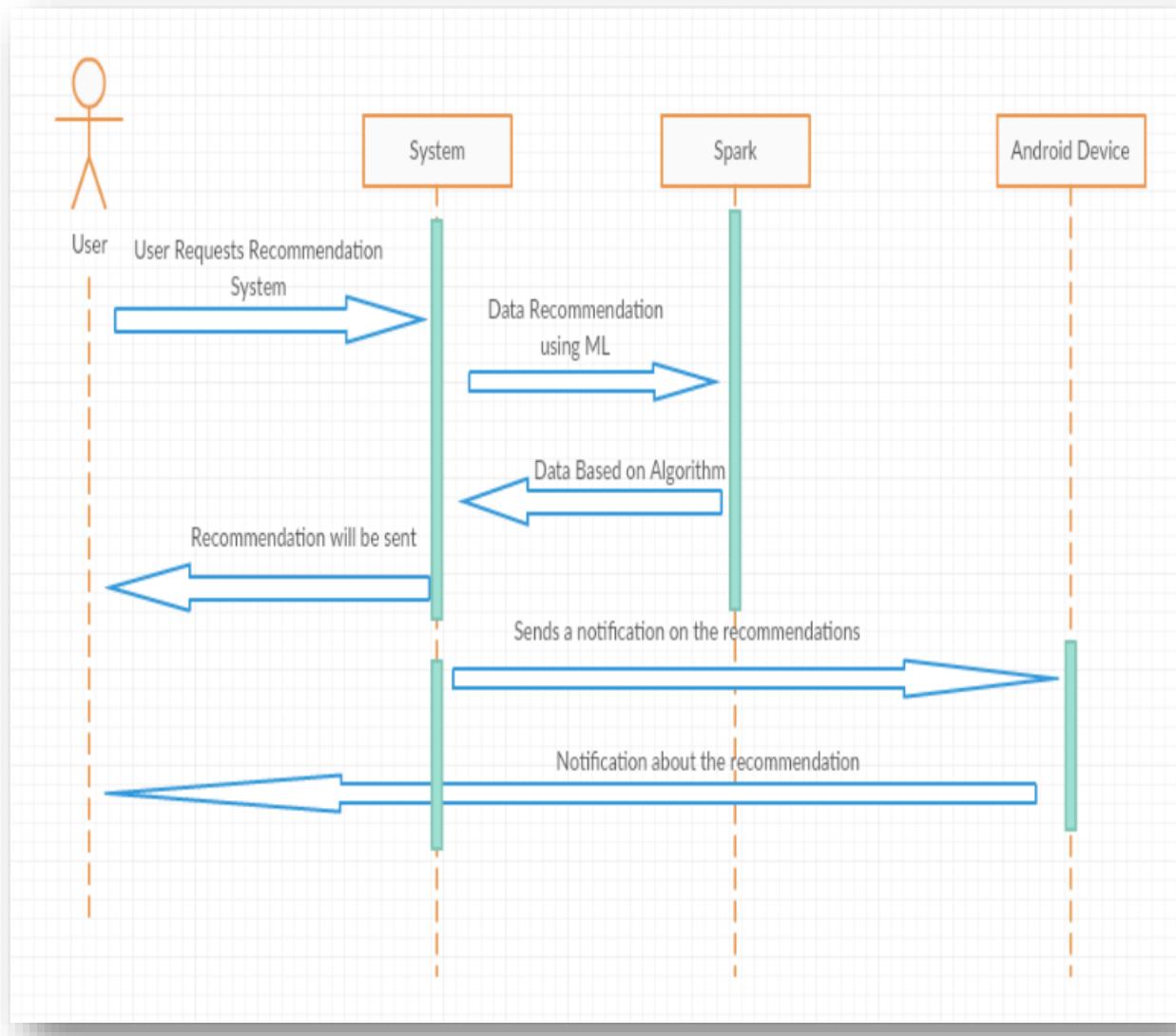
7 IMPLEMENTATION

7.1 ARCHITECTURE & UML DIAGRAMS:

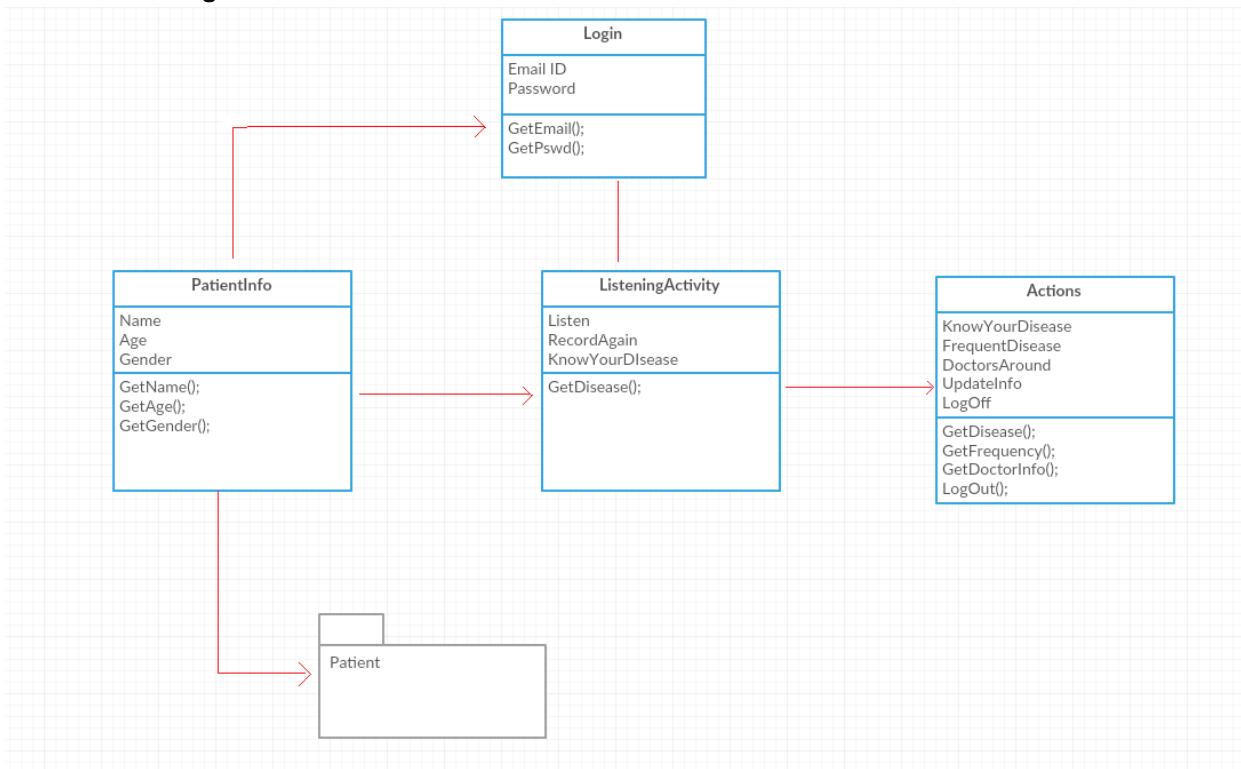


In traditional way, diagnostic test will be done first and then a drug will be used to a person where if he responds to it, then doctor will recommend to continue that drug A for better outcome, if by any reason the patient will not respond to that drug A, then doctor will recommend drug B. At-last if patient shows some severe side effects even after using drug B, then doctor will recommend drug C which is a long and time taking process. This can be changed by the above proposed approach.

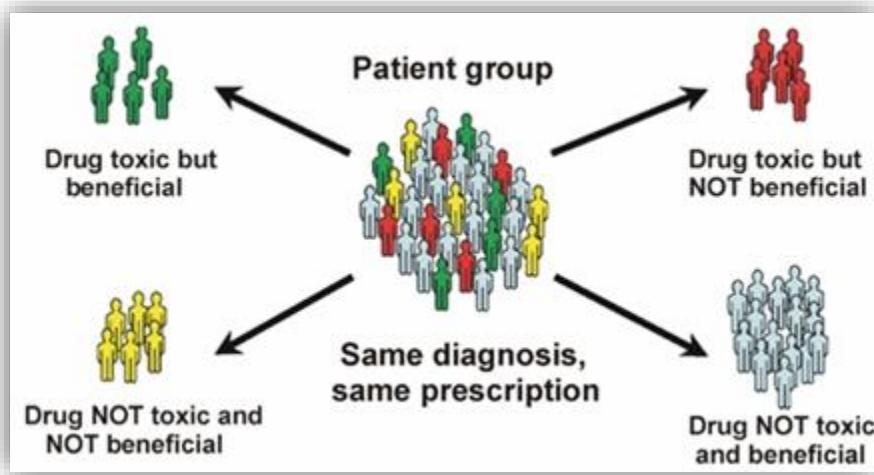
7.1.1 Sequence Diagram:



7.1.2 Class Diagram:



7.2 WORKFLOW:



Here we have a data set with group of people. Later this machine should segregate the data to groups like drug toxic but beneficial for some time, Drug toxic but NOT beneficial, Drug Not Toxic and Not Beneficial and Drug Not Toxic and Beneficial (this group will be ideal for data collection). Based on this we will give rating to the drugs for that particular group and later we will use that data for drug recommendation.

7.3 EXISTING SERVICES:

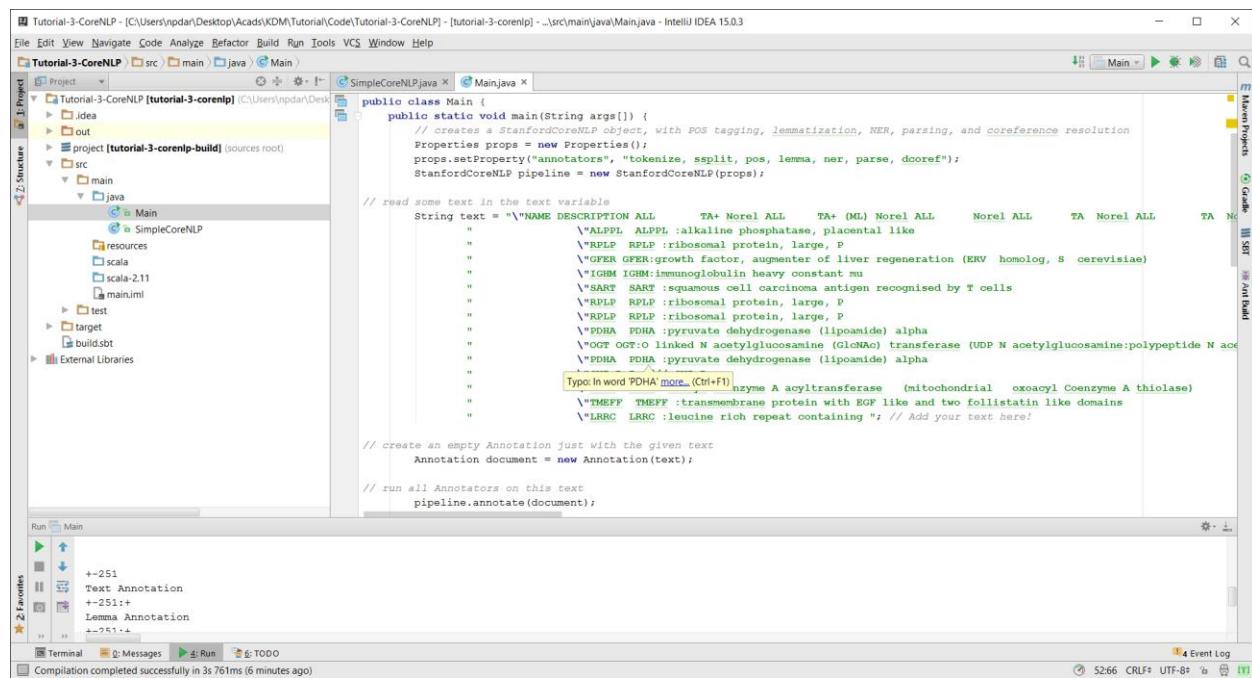
7.3.1 First Increment:

We have used the following Existing services:

Collaborative filtering (CF) is a technique that we used to filter the information and produce an output Dataset. Collaborative filtering methods have been used on the input data which include the symptoms and identifies the type of diseases. Later it is used to analyze the drugs that are prescribed for such kinds of diseases based on the given criteria.

Natural Language Processing (NLP) is the technique that we have used to processes the input given by the User. Here the NLP takes the input as a symptoms and it processes the data to enable computer to derive meaning.

NLP Data provided:



The screenshot shows the IntelliJ IDEA 15.0.3 interface with the following details:

- Project Structure:** Shows the project structure for "Tutorial-3-CoreNLP [tutorial-3-corenlp]" located at "C:\Users\ndpar\Desktop\Acads\KDM\Tutorial\Code\Tutorial-3-CoreNLP". It includes a .idea folder, out, src (with main and java subfolders), test, target, build.sbt, and External Libraries.
- Main.java:** The active file contains Java code for a Main class. The code imports various NLP-related classes and defines a main method that reads text from a string variable and processes it through a StanfordCoreNLP pipeline.
- Annotations:** A tooltip appears over the word "PDHA", indicating it is annotated as "Type: In word PDHA more_ (Ctrl+Shift+F) enzyme A acyltransferase (mitochondrial oxaacyl Coenzyme A thiolase)".
- Run Tab:** Shows a run configuration for "Main" with a "Run" button and a dropdown menu showing "251", "Text Annotation", "251+", "Lemma Annotation", and "251+-".
- Bottom Status Bar:** Displays "Compilation completed successfully in 3s 761ms (6 minutes ago)" and "52:66 CRLF: UTF-8 'a' Event Log".

Project Report Increment 2

Parse tree:

The screenshot shows the IntelliJ IDEA interface with the following details:

- Project Structure:** The project is named "Tutorial-3-CoreNLP". It contains a "src" directory with a "main" package, which in turn contains a "java" directory and a "Main.java" file.
- Main.java Code:** The code prints the parse tree and semantic graph for a sentence. It uses `System.out.println` to output the results.

```
System.out.println("\n\n");
}

// this is the parse tree of the current sentence
Tree tree = sentence.get(TreeCoreAnnotations.TreeAnnotation.class);
System.out.println(tree);
// this is the Stanford dependency graph of the current sentence
SemanticGraph dependencies = sentence.get(SemanticGraphCoreAnnotations.CollapsedCCProcessedDependenciesAnnotation.class);
System.out.println(dependencies.toString());
```

Run Tab: The "Main" configuration is selected.

Left Sidebar: Shows the "Structure" view with a tree diagram of the code's dependencies.

Bottom Status Bar: Shows the message "Compilation completed successfully in 3s 761ms (8 minutes ago)".

Process execution:

The screenshot shows the IntelliJ IDEA interface with the following details:

- Title Bar:** Tutorial-3-CoreNLP - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\Code\Tutorial-3-CoreNLP] - [tutorial-3-corenlp] - ...src\main\java\Main.java - IntelliJ IDEA 15.0.3
- Menu Bar:** File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
- Toolbar:** Main
- Project Tool Window:** Shows the project structure for "Tutorial-3-CoreNLP [tutorial-3-corenlp]" with "src" and "out" directories.
- Code Editor:** Displays the Main.java file with the following code:

```

        System.out.println("\n\n");
    }

    // this is the parse tree of the current sentence
    Tree tree = sentence.get(TreeCoreAnnotations.TreeAnnotation.class);
    System.out.println(tree);
    // this is the Stanford dependency graph of the current sentence
    SemanticGraph dependencies = sentence.get(SemanticGraphCoreAnnotations.CollapsedCCProcessedDependenciesAnnotation.class);
    System.out.println(dependencies.toString());
}

```
- Run Tool Window:** Shows the output of the application. The output window displays the generated parse tree and dependency graph for the sentence "Coenzyme I in sentence 1".
- Status Bar:** Compilation completed successfully in 3s 761ms (9 minutes ago)

7.3.2 Second Increment:

WordNet on top TFIDF words:

```

String start = "medicine";
String end = "drug";
pos = wordnet.getBestPos(start);

// Wordnet can find relationships between words
System.out.println("\n\nRelationship between: " + start + " and " + end);
float dist = wordnet.getDistance(start,end, pos);
String[] parents = wordnet.getCommonParents(start, end, pos);
System.out.println(start + " and " + end + " are related by a distance of: " + dist);

// These words have common parents (hyponyms in this case)
System.out.println("Common parents: ");
if (parents != null) {
    for (int i = 0; i < parents.length; i++) {
        System.out.println(parents[i]);
    }
}

// System.out.println("\n\nHyponym Tree for " + start);
// int[] ids = wordnet.getSenseIds(start,wordnet.NOUN);
// wordnet.printHyponymTree(ids[0]);

```

```

"C:\Program Files\Java\jdk1.8.0_65\bin\java" ...
Finding parts of speech for page.
page
pagev

Definitions for patent:
patentobtain a patent for
patentgrant rights to; grant a patent for
patentmake open to sight or notice

Synonyms for next (pos: a)
next - adjacent
next - close
next - following
next - future
next - incoming
next - side by side
next - succeeding

Synonyms for next (pos: r)

Synonyms for page (pos: n)
page - Page
page - Sir Frederick Handley Page
page - Thomas Nelson Page
page - arouser
page - baggageman

```

Project Report Increment 2

Screenshot of IntelliJ IDEA 15.0.3 showing the WordNetMain application running. The terminal shows the application has completed successfully.

```
String start = "medicine";
String end = "drug";
pos = wordnet.getBestPos(start);

// Wordnet can find relationships between words
```

Synonyms for page (pos: v)

- page - assist
- page - beaver
- page - beaver away
- page - beckon
- page - boondoggle
- page - busy
- page - buzz
- page - carpenter
- page - clerk
- page - collaborate

Synonyms for doctor (pos: n)

- doctor - Arabist
- doctor - Cabalist
- doctor - Church Father
- doctor - Doctor
- doctor - Doctor of the Church
- doctor - Dr
- doctor - Father
- doctor - Father of the Church
- doctor - GP
- doctor - Islamist

Synonyms for doctor (pos: v)

- doctor - administer
- doctor - advance
- doctor - aid
- doctor - alleviate
- doctor - analyse

Screenshot of IntelliJ IDEA 15.0.3 showing the WordNetMain application running. The terminal shows the application has completed successfully.

```
String start = "medicine";
String end = "drug";
pos = wordnet.getBestPos(start);

// Wordnet can find relationships between words
```

Synonyms for currently (pos: r)

- currently - presently

Synonyms for change (pos: n)

- change - Coriolis effect
- change - Maundy money
- change - about-face
- change - acceleration
- change - accenting
- change - accentuation
- change - accessory
- change - accident
- change - accompaniment
- change - accomplishment

Synonyms for change (pos: v)

- change - Americanise
- change - Americanize
- change - Europeanise
- change - Europeanize
- change - Frenchify
- change - Islamise
- change - Islamize
- change - accelerate
- change - accommodate
- change - accompany

Synonyms for rising (pos: a)

- rising - acclivitous

Project Report Increment 2

SparkOpenE - [C:\Users\Npda\Desktop\Acads\KDM\Tutorial\CS5560 - Tutorial 7 Source Code\SparkOpenE] - [sparkopenie] - ..\src\main\scala\wordnet\WordNetMain.java - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkOpenE > src > main > scala > wordnet > WordNetMain

Project Structure

Run WordNetMain

```

String start = "medicine";
String end = "drug";
pos = wordnet.getBestPos(start);

// Wordnet can find relationships between words

```

Synonyms for brand (pos: v)
brand - address
brand - asterisk
brand - badge
brand - bespot
brand - brandmark
brand - calibrate
brand - call
brand - code
brand - crisscross
brand - delineate

Synonyms for mg (pos: n)
mg - Al
mg - Am
mg - Ba
mg - Be
mg - Bi
mg - Bk
mg - Ca
mg - Cd
mg - Ce
mg - Cf

Synonyms for percentage (pos: n)
percentage - absentee rate
percentage - accounts receivable
percentage - allocation
percentage - allotment

Event Log

Compilation completed successfully in 6s 971ms (2 minutes ago)

Ask me anything

SparkOpenE - [C:\Users\Npda\Desktop\Acads\KDM\Tutorial\CS5560 - Tutorial 7 Source Code\SparkOpenE] - [sparkopenie] - ..\src\main\scala\wordnet\WordNetMain.java - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkOpenE > src > main > scala > wordnet > WordNetMain

Project Structure

Run WordNetMain

```

String start = "medicine";
String end = "drug";
pos = wordnet.getBestPos(start);

// Wordnet can find relationships between words

```

mg - Cf

Synonyms for percentage (pos: n)
percentage - absentee rate
percentage - accounts receivable
percentage - allocation
percentage - allotment
percentage - allowance
percentage - amount
percentage - amount of money
percentage - batting average
percentage - capital
percentage - case-fatality proportion

Hyponyms for million:

Hypernyms for million:
large integer
large indefinite quantity
large indefinite amount

Relationship between: medicine and drug
medicine and drug are related by a distance of: 0.16666667
Common parents:
entity

Process finished with exit code 0

Event Log

Compilation completed successfully in 6s 971ms (2 minutes ago)

Ask me anything

Word to Vector using MLLIB:

Word 2 Vec takes input as a text/JSON corpus and gives output in the form of feature vectors for the words in that text corpus. Feature vector is a vector that is formed after doing Lemmatization, stop word removal. Word2vec trains words against other words that neighbor them in the input corpus. Similarities are done between words based on the associations.

We gave medicaments.json as input file which contains data about existing drugs and their composition, manufactured date, number of them available etc.

```

Spark-W2Vec - [C:\Users\lakshmi\Desktop\CS5560-T6-Source Code\CS5560 - Tutorial 6 Source Code\Spark-W2Vec] - [spark-w2vec] - ...src\main\scala\ml\pipeline\SparkW2V\MLscala - IntelliJ IDEA 2016.1.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
File Project Packages Project Files SparkW2Vec src main scala mlpipeline SparkW2VMLscala
Project Structure
Run SparkW2VML
Java Enterprise Run TODO Terminal
Compilation completed successfully in 6s 768ms (7 minutes ago)
27881:1 CRLF: UTF-8: 3:14 PM 7/8/2016

```

```

// Turn off Info Logger for Console
logger.getLogger("org").setLevel(Level.OFF);
logger.getLogger("akka").setLevel(Level.OFF);

// Read the file into RDD[String]
val input = sc.textFile("data/medicaments").map(line => {
    // Getting Isbn from the file
    val lemma = TypoInWord('medicaments' more_ (Ctrl+F))
    (line, lemma)
})

```

Project Report Increment 2

The screenshot shows the IntelliJ IDEA interface with the following details:

- Project Structure:** The project is named "Spark-W2Vec [spark-w2vec]". It contains a "src" directory with "main" and "mlpipeline" sub-directories. "mlpipeline" contains Scala files like "SparkW2VML.scala", "TFIDF.scala", and "SparkW2V.scala".
- Code Editor:** The main editor window displays the "SparkW2VML.scala" file. The code initializes a SparkSession, reads a text file of medications, and performs NLP processing (lemmatization, tokenization, stop word removal) to create a DataFrame.
- Run Tab:** The "Run" tab shows a successful run of the "SparkW2VML" application.
- Bottom Bar:** Shows various tool icons and the status bar indicating "Compilation completed successfully in 6s 768ms (9 minutes ago)".

Output snippet is shown below

```
bloc :  
[0,8,0.999991693342083]  
[séborrhéique,0.9999624167176876]  
[trafic,0.999951455955272]  
  
power :  
[10/07/2013,0.9995441497647819]  
[2003-08-18,0.9993619630221745]  
[sous-cutanées,0.9993440385325913]  
  
décurarisation :  
[nanocolloïdes,0.9999969329010768]  
[marcaine,0.9999789195759184]  
[954-5,0.9999456291830718]  
  
neuromusculaire :  
[64975424,0.9999907636165579]  
[18,2,0.9999839484435472]  
[suspicion,0.9999696840965253]  
  
induit :  
[prévention,0.9999922947274583]  
[theostat,0.9998890689560397]  
[rhumatismes,0.9998063200484436]  
  
rocuronium :  
[cytotec,0.9999636647599903]  
[213-5,0.9998232878230584]  
[fluocyne,0.9998151128176668]  
  
cymevan :  
[aas,0.9999503875533897]  
[membrane,0.9999468208849455]  
[antigène,0.9999024447530916]
```

K-MEANS:

In K Means we can cluster the whole bunch of documents and extract some most common key words to represent the topic of the document.

```
Reading POS tagger model from edu/stanford/nlp/models/pos-tagger/english-left3words/english-left3words-distsim.tagger ... done [2.8 sec].
root
|--- location: string (nullable = true)
|--- docs: string (nullable = true)
|--- rawTokens: array (nullable = true)
|   |--- element: string (containsNull = true)
|--- tokens: array (nullable = true)
|   |--- element: string (containsNull = true)
|--- features: vector (nullable = true)

()
[file:///E:/#UMKC/KDM/Tutorial/9/CS5560 - Tutorial 9 Source Code/Spark_KMeans_FV/data/drugs/med.Insulin/1400.txt,RDF datum ospa lipoprotein Resource URI http wifo informatik unus

Corpus summary:
  Training set size: 12 documents
  Vocabulary size: 968 terms
  Preprocessing time: 27.49186883 sec

16/07/06 21:18:25 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
16/07/06 21:18:25 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
Finished training KMeans model. Summary:
  Training time: 1.951532782 sec

Process finished with exit code 0
```

LDA:

```

SparkLDA - [C:\Users\Npdar\Desktop\Acads\CS5560 - Tutorial 8 Source Code\SparkLDA] - [sparkLda] - ...data\ner\domainBasedWords - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SparkLDA data ner domainBasedWords
Project SparkLDA [sparkLda] C:\Users\Npdar\Desktop\Acads\CS5560
  .idea
  data
    20_news_group
      sci_crypt.txt
      sci_electronics.txt
      sci_med.txt
      sci_space.txt
    ner
      domainBasedWords
      sample
      Results.txt
  Run SparkLDAMain
  18 5.325394948025887E-4
  2016 5.325394947943757E-4
  fewer 5.201650911160667E-4
  expiry 5.201650911038421E-4
  ; 5.201650910887035E-4
  lastly 5.20165091074147E-4
  justify 5.201650910729724E-4
  period 5.201650910465048E-4
  capture 5.201650910420657E-4
  y 5.201650910376502E-4
  ' 5.201650910265634E-4
  lifesaving 5.201650910202453E-4
  competitor 5.20165091014611E-4
  al 5.201650910118556E-4
  2.6 5.201650910107151E-4
  hinder 5.201650910090387E-4
  develop 5.20165090998469E-4
  generic 5.2016509095979478E-4
  70 5.20165090977437E-4

  Process finished with exit code 0
  
```

Terminal Run TODO

All files are up-to-date (a minute ago)

Ask me anything

Event Log 149851 CRLF+ UTF-8 405 PM 7/8/2016

```

SparkLDA - [C:\Users\Npdar\Desktop\Acads\CS5560 - Tutorial 8 Source Code\SparkLDA] - [sparkLda] - ...data\Results.txt - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SparkLDA data Results.txt
Project SparkLDA [sparkLda] C:\Users\Npdar\Desktop\Acads\CS5560
  .idea
  data
    20_news_group
      sci_crypt.txt
      sci_electronics.txt
      sci_med.txt
      sci_space.txt
    ner
      domainBasedWords
      sample
      Results.txt
    project [sparkLda-build] (sources root)
    spark-warehouse
    src
      main
        java
  Run SparkLDAMain
  Corpus summary:
    Training set size: 97 documents
    Vocabulary size: 741 terms
    Training set size: 1494 tokens
    Preprocessing time: 15.763828637 sec

  Finished training LDA model. Summary:
    Training time: 11.916366596 sec
    Training data average log likelihood: -110.6917543173391

  20 topics:
  TOPIC_0_,;0.08755849653882213
  TOPIC_0_,;0.056209948269393406
  TOPIC_0_,drug,0.04620890538997851
  TOPIC_0_,price,0.012336901228243068
  TOPIC_0_,say,0.01070585550805288
  TOPIC_0_,company,0.009146675743181202
  TOPIC_0_,'',0.008824877931782904
  TOPIC_0_,'',0.008313260817210013
  TOPIC_0_,';,0.007579735772978709
  cancer,0.0075009512730609545
  --,0.0070139514512827
  spending,0.0069025923049940326

  Training data average log likelihood: -110.6917543173391

  20 topics:
  TOPIC_0_
  drug,0.08755849653882213
  price,0.056209948269393406
  say,0.01070585550805288
  company,0.009146675743181202
  '',0.008824877931782904
  --,0.008313260817210013
  $,0.007579735772978709
  cancer,0.0075009512730609545
  --,0.0070139514512827
  spending,0.0069025923049940326
  
```

Terminal Run TODO

All files are up-to-date (2 minutes ago)

Ask me anything

Event Log 15:33 CRLF+ UTF-8 405 PM 7/8/2016

Named Entity Relation results based on the domain based words:

```

**
** Created by Mayanka on 29-Jun-16.
**
object SparkNER {
  def main(args: Array[String]) {
    System.setProperty("hadoop.home.dir", "C:\\Users\\npdar\\Desktop\\Acads\\BigDataAnalytics\\winutils")
    val conf = new SparkConf().setAppName("NERTrain").setMaster("local[*]").set("spark.driver.memory", "4g").set("spark.executor.memory", "4g")
    val sc = new SparkContext(conf)
  }
}

```

```

**
** Created by Mayanka on 29-Jun-16.
**
object SparkNER {
  def main(args: Array[String]) {
    System.setProperty("hadoop.home.dir", "C:\\Users\\npdar\\Desktop\\Acads\\BigDataAnalytics\\winutils")
    val conf = new SparkConf().setAppName("NERTrain").setMaster("local[*]").set("spark.driver.memory", "4g").set("spark.executor.memory", "4g")
    val sc = new SparkContext(conf)
  }
}

```

Project Report Increment 2

SparkLDA - [C:\Users\Npdar\Desktop\Acads\CS5560 - Tutorial 8 Source Code\SparkLDA] - [sparkida] - ...\\src\\main\\scala\\ner\\SparkNER.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkLDA > src > main > scala > ner > SparkNER.scala

```

object SparkNER {
    def main(args: Array[String]) {
        System.setProperty("hadoop.home.dir", "C:\\Users\\Npdar\\Desktop\\Acads\\BigDataAnalytics\\winutils")
        val conf = new SparkConf().setAppName("NERTrain").setMaster("local[*]").set("spark.driver.memory", "4g").set("spark.executor.memory", "4g")
        val sc = new SparkContext(conf)
    }
}

```

Run: SparkNER

(this, o)
 (will, o)
 (actually, o)
 (increase, o)
 (spending, ' ', o o)
 (say, o)
 (Caroline, PERSON)
 (Pearson, PERSON)
 (of, o)
 (consult, o)
 (firm, o)
 (Avalere, o)
 (in, o)
 (Washington, LOCATION o)
 (D.C., , LOCATION o)
 (which, o)
 (be, o)
 (in, o)
 (the, o)
 (process, o)
 (of, o)
 (analyze, o)
 (the, o)
 (financial, o)
 (impact, o)
 (of, o)
 (the, o)
 (meanwhile, o)

Terminal Run TODO Event Log

All files are up-to-date (2 minutes ago)

145 chars, 9 lines 29422 CRLF+ UTF-8 4:12 PM 7/8/2016

SparkLDA - [C:\Users\Npdar\Desktop\Acads\CS5560 - Tutorial 8 Source Code\SparkLDA] - [sparkida] - ...\\src\\main\\scala\\ner\\SparkNER.scala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkLDA > src > main > scala > ner > SparkNER.scala

```

object SparkNER {
    def main(args: Array[String]) {
        System.setProperty("hadoop.home.dir", "C:\\Users\\Npdar\\Desktop\\Acads\\BigDataAnalytics\\winutils")
        val conf = new SparkConf().setAppName("NERTrain").setMaster("local[*]").set("spark.driver.memory", "4g").set("spark.executor.memory", "4g")
        val sc = new SparkContext(conf)
    }
}

```

Run: SparkNER

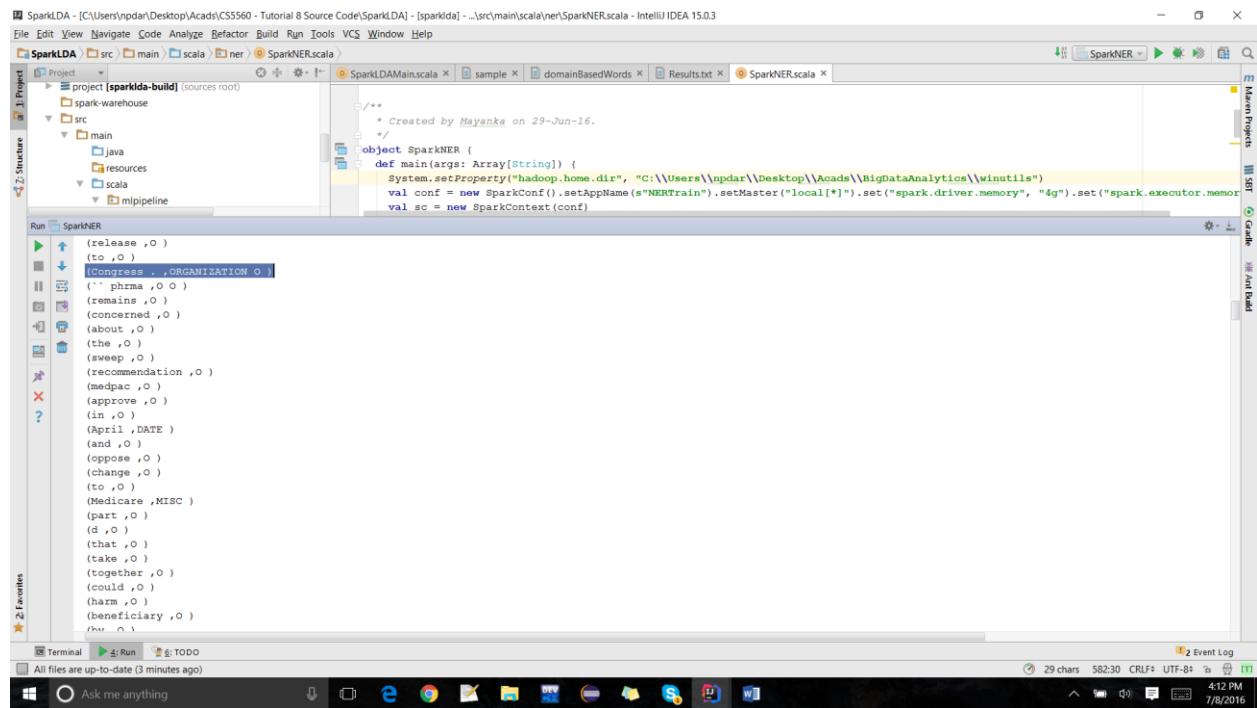
(factor, o)
 (help, o)
 (drive, CS)
 (spending, o)
 (in, o)
 (the, o)
 (program, o)
 (to, o)
 (\$ 73, MONEY MONEY)
 (billion, NUMBER)
 (in, o)
 (2014, DATE)
 (-, o)
 (and, o)
 (more, o)
 (high-cost, o)
 (drug, o)
 (will, o)
 (hit, o)
 (the, o)
 (market, o)
 (soon, , o o)
 (without, o)
 (action, , o o)
 (the, o)
 (rise, CS)
 (cost, o)
 (could, o)

Terminal Run TODO Event Log

All files are up-to-date (3 minutes ago)

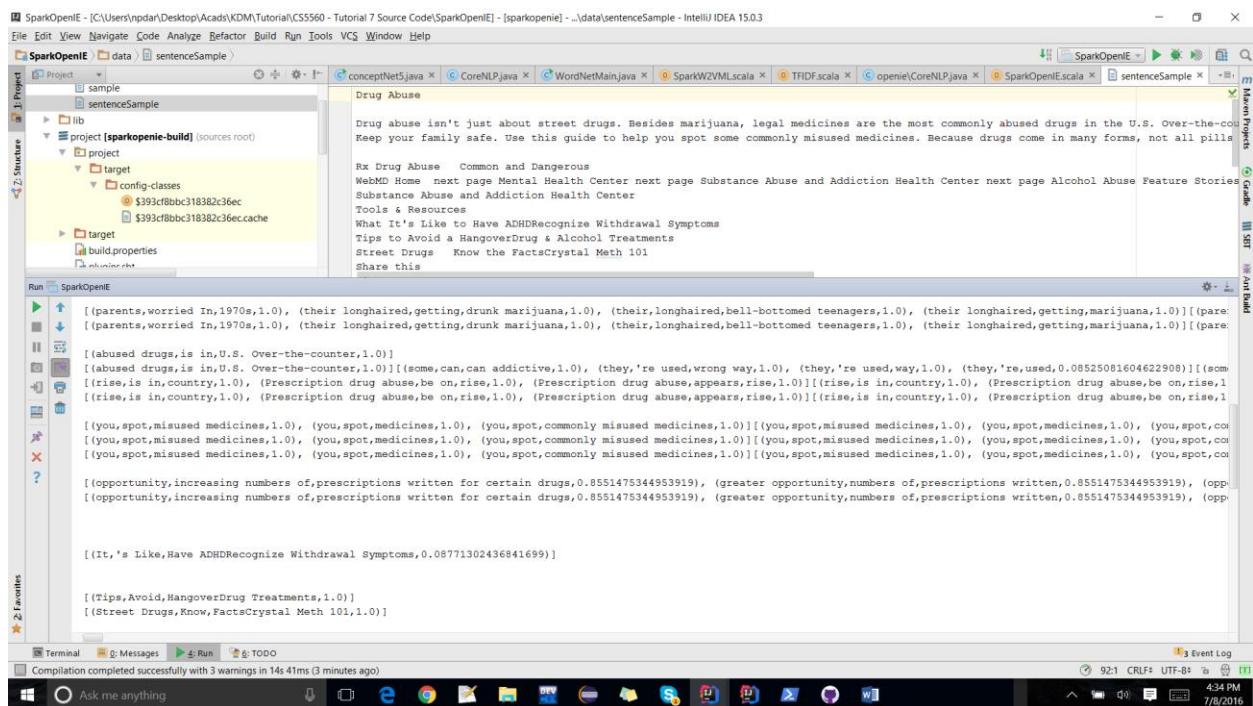
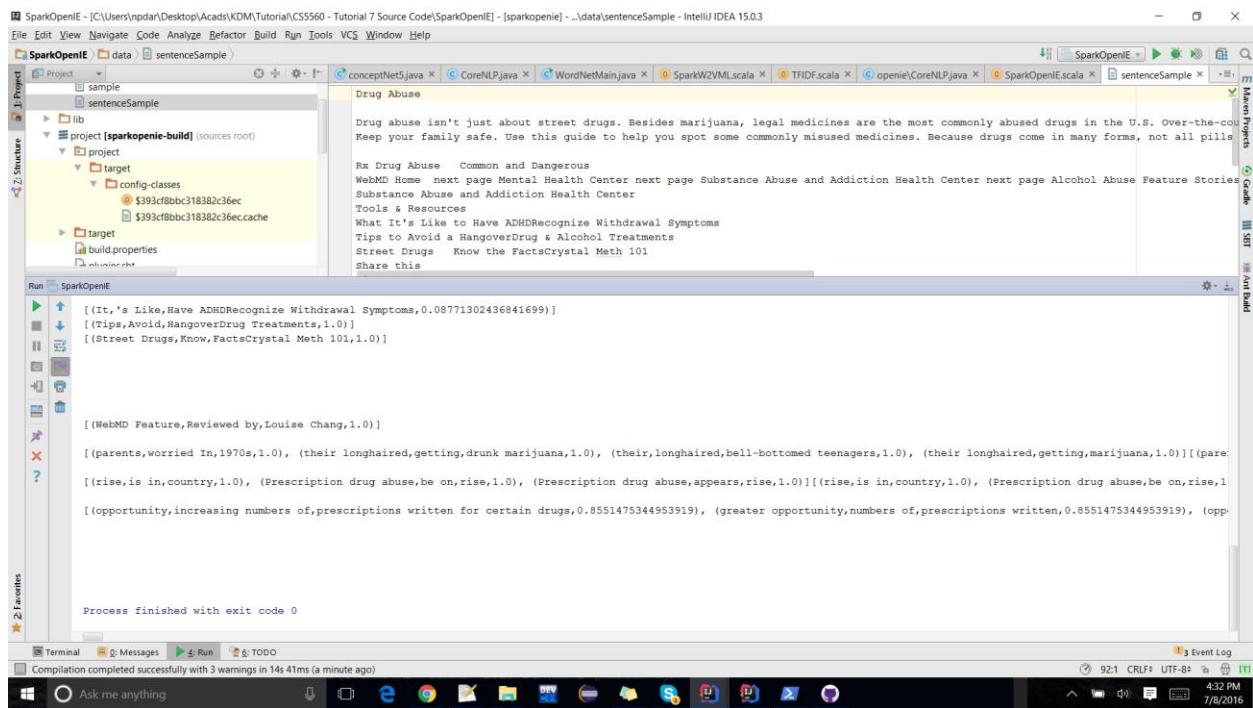
39 chars, 2 lines 48419 CRLF+ UTF-8 4:12 PM 7/8/2016

Project Report Increment 2

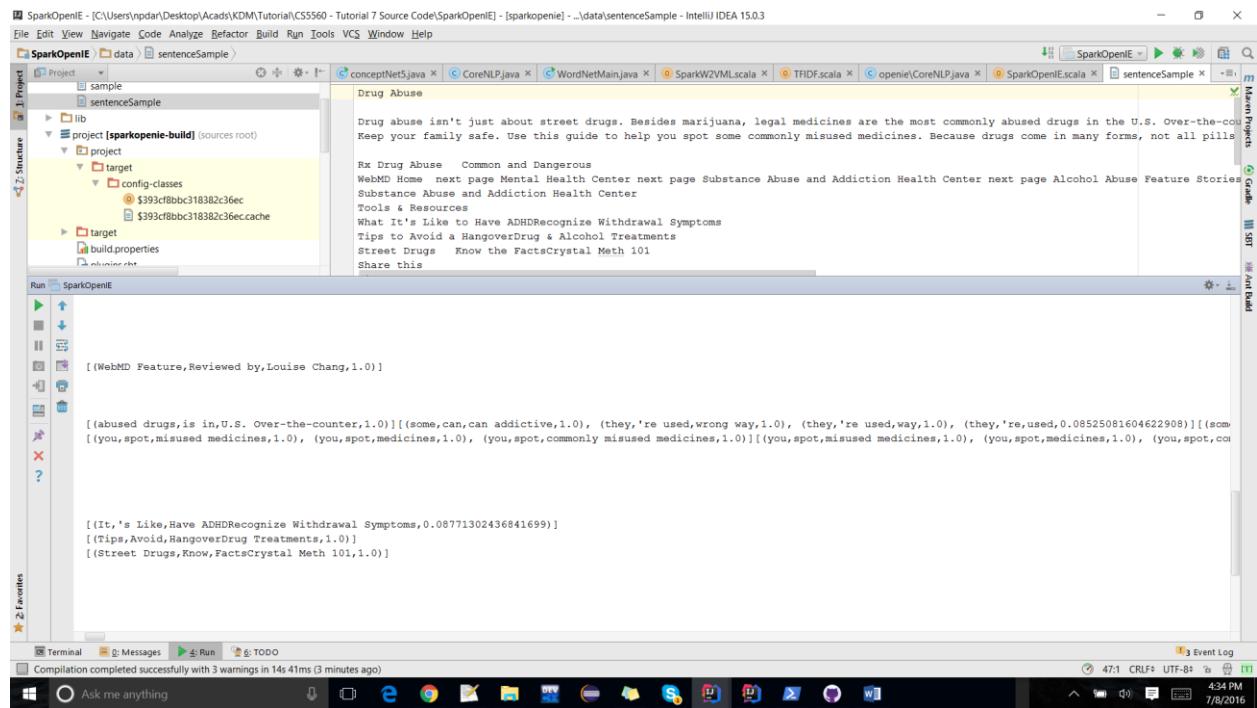


Project Report Increment 2

OPEN IE:



Project Report Increment 2



Comparison:

Both K-Means and LDA are unsupervised learning, also both will have parameters like K value, number of clusters and number of topics. For suppose if we are K topics in set of N documents, K-Means will N number of documents into K disjoint clusters (topics). But in LDA, it assigns a document of many different topics which are mixed, because of which we are getting more than one topic for each document. Therefore, LDA will be more realistic than K-Means in topic assignment.

7.4 FEATURE IMPLEMENTED:

TF-IDF: We have used TF-IDF to extract the top drugs that can be recommended to the user based on the symptoms and the criteria of the overall system. It reflects how important the word is to a document or to a dataset. Based on the provided statistics, we consider the similarities with the training data and extract the relevant top 10 drugs that can be recommended.

TF-IDF – Dataset:

```

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
Spark-TF-IDF Article.txt IntelliJ IDEA 15.0.3
Spark-TF-IDF Article.txt
Project Spark-TF-IDF Article.txt
1: Project
2: Favorites
3: Structure
4: Article.txt
IRAK interleukin receptor associated kinase
UBE A Ubiquitin conjugating enzyme E A (RAD homolog)
PRKCQ protein kinase C, iota
COX I cytochrome c oxidase subunit IV isoform
PRKCH protein kinase C, eta
UBE I Ubiquitin conjugating enzyme E I (UBC homolog, yeast)
PRKCG protein kinase C, gamma
UBE H Ubiquitin conjugating enzyme E H (UBC homolog, yeast)
PRKCE protein kinase C, epsilon
UBE C Ubiquitin conjugating enzyme E C
PRKCD protein kinase C, delta
UBE B Ubiquitin conjugating enzyme E B (RAD homolog)
ENO B:
PROC protein C (inactivator of coagulation factors Va and VIIIa)
PRKD protein kinase D
UBE N Ubiquitin conjugating enzyme E N (UBC homolog, yeast)
PRKCQ protein kinase C, theta
GHRH growth hormone releasing hormone
ARCN arachin
COMMD copper metabolism (Murr ) domain containing
NGFR nerve growth factor receptor (TNFR superfamily, member )
PRNP prion protein (P ) (Creutzfeldt Jakob disease, Gerstmann Strausler Scheinker syndrome, fatal familial insomnia)
AOC amine oxidase, copper containing (retina specific)
UBE S Ubiquitin conjugating enzyme E S
MATR matrin
AOC amine oxidase, copper containing (vascular adhesion protein )
MBL mannose binding lectin (protein C) , soluble (spastic defect)
PRK2 PRK2:protein kinase C, zeta
UBE A Ubiquitin protein ligase E A (human papilloma virus E associated protein, Angelman syndrome)
TRA A:
GPR G protein coupled receptor
PRKDC PRKDC:protein kinase, DNA activated, catalytic polypeptide
HCFC host cell factor C (VP accessory protein)
AKAP AKAP :A kinase (PRKA) anchor protein (yotiao)
UBE C Ubiquitin protein ligase E C
SMA // SMA na
NGFB nerve growth factor, beta polypeptide
GPR G protein coupled receptor

```

Project Report Increment 2

Output:

```

Spark-TF-IDF - [C:\Users\ndpar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF] - [spark-tf-idf - ...src\main\scala\TFIDF_Main.scala - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
Spark-TF-IDF src main scala TFIDF_Main.scala
Project Structure Favorites Run Terminal Event Log
Spark-TF-IDF [spark-tf-idf] (C:\Users\ndpar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF) - [spark-tf-idf - ...src\main\scala\TFIDF_Main.scala - IntelliJ IDEA 15.0.3
src project [spark-tf-idf-build] (sources root)
  main
    java
    resources
  Article.txt
  TFIDF_Main.scala
  Run TFIDF_Main
  Run Terminal
  Run Event Log
  Run TODO
  All files are up-to-date (a minute ago)
  5846:80 CRLF: UTF-8: 's' 4 Event Log
  
```

```

object TFIDF_Main {
  def main(args: Array[String]): Unit = {
    val sc = SparkContext.getOrCreate()
    val rdd = sc.textFile("Article.txt")
    val words = rdd.flatMap(_.split(" "))
    val wordCount = words.count()
    val wordRdd = words.map((_, 1))
    val wordCountRdd = wordRdd.reduceByKey(_ + _)
    val wordCountMap = wordCountRdd.collect().toMap
    val tfidfData = wordRdd.map { wordAndOne =>
      val word = wordAndOne._1
      val count = wordAndOne._2
      val tf = count / wordCount
      val idf = Math.log(wordCount / wordCountMap.get(word))
      (word, tf * idf)
    }
    tfidfData.foreach(f => println(f))
    val mapp = sc.broadcast(tfidfData.zip(tfidfData.values).map { case (tfidf, value) => (tfidf, value) })
    mapp.value.foreach(f => println(f))
  }
}
  
```

Top 10 terms:

```

Spark-TF-IDF - [C:\Users\ndpar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF] - [spark-tf-idf - ...src\main\scala\TFIDF_Main.scala - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
Spark-TF-IDF [spark-tf-idf] (C:\Users\ndpar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF) - [spark-tf-idf - ...src\main\scala\TFIDF_Main.scala - IntelliJ IDEA 15.0.3
src project [spark-tf-idf-build] (sources root)
  main
    java
    resources
  Article.txt
  TFIDF_Main.scala
  Run TFIDF_Main
  Run Terminal
  Run Event Log
  Run TODO
  All files are up-to-date (a minute ago)
  5846:80 CRLF: UTF-8: 's' 4 Event Log
  
```

```

16/06/24 22:07:38 INFO TaskSetManager: Finished task 0.0 in stage 8.0 (TID 14) in 24 ms on localhost (1/1)
16/06/24 22:07:38 INFO TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool
16/06/24 22:07:38 INFO DAGScheduler: ResultStage 8 (take at TFIDF_Main.scala:56) finished in 0.027 s
16/06/24 22:07:38 INFO DAGScheduler: Job 5 finished: take at TFIDF_Main.scala:56, took 0.516311 s
16/06/24 22:07:38 INFO SparkContext: Invoking stop() for shutdown hook
  (more...) at87.2934288651337
  (MDL, 63.4961300837534)
  (TA, 47.6145975628002)
  (MDL_ALL, 47.6145975628002)
  (AML, 47.6145975628002)
  ((PK), 31.7430650418668)
  ((TA+), 31.7430650418668)
  (neurokinin, 23.8072987814001)
  (Norel, 15.8715325209334)
  (DB, 15.8715325209334)
  (RBMY, 15.8715325209334)
  (Relap, AML, 15.8715325209334)
  (GNTL, 15.8715325209334)
  (NOMO, AML, 15.8715325209334)
  (CTAG, 15.8715325209334)
  (CC, 15.8715325209334)
  (DAZ, 15.8715325209334)
  (AFB, 15.8715325209334)
  (KB, 15.8715325209334)
16/06/24 22:07:38 INFO SparkUI: Stopped Spark web UI at http://192.168.6.1:4040
16/06/24 22:07:38 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/06/24 22:07:38 INFO MemoryStore: MemoryStore cleared
  
```

Process Success:

```

// Spark-TF-IDF - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF] - [spark-tf-idf] - ...\\src\\main\\scala\\TFIDF_Main.scala - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
Spark-TF-IDF src main scala > TFIDF_Main.scala
Project Structure Run Terminal Event Log
  Project Spark-TF-IDF [spark-tf-idf] C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF
    src
      project [spark-tf-idf-build] sources root
        main
          java
          resources
    Run TFIDF_Main
      Article.txt
        val tfidf = new TFIDF()
        tfidf.foreach(f => println(f))
        val tfidfData = tfidfindex.zip(tfidfvalues)
        var hm = new HashMap[String, Double]
        tfidfData.collect().foreach(f => {
          hm += f._1 -> f._2.toDouble
        })
        val mapp = sc.broadcast(hm)
      Output
        AML,47.6145975628002
        ((PK),31.7430650418668
        (TA+,31.7430650418668
        (neurokinin,23.8072987814001
        (Norel AML,15.8715325209334
        (DS,15.8715325209334
        (RBMY,15.8715325209334
        (Relap AML,15.8715325209334
        (GNTL,15.8715325209334
        (NOMO,15.8715325209334
        (MG_AML,15.8715325209334
        (CTAG,15.8715325209334
        (CC,15.8715325209334
        (DAZ,15.8715325209334
        (AFB,15.8715325209334
        (KB,15.8715325209334
        16/06/24 22:07:38 INFO SparkUI: Stopped Spark web UI at http://192.168.6.1:4040
        16/06/24 22:07:38 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
        16/06/24 22:07:38 INFO MemoryStore: MemoryStore cleared
        16/06/24 22:07:38 INFO BlockManager: BlockManager stopped
        16/06/24 22:07:38 INFO BlockManagerMaster: BlockManagerMaster stopped
        16/06/24 22:07:38 INFO SparkContext: Successfully stopped SparkContext
        16/06/24 22:07:38 INFO ShutdownHookManager: Shutdown hook called
        16/06/24 22:07:38 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
        16/06/24 22:07:38 INFO ShutdownHookManager: Deleting directory C:\Users\ndar\AppData\Local\Temp\spark-b14a7ba5-7b9b-40a0-ad8c-0ab98ac9dfaf
        Process finished with exit code 0
  Favorites Terminal Run Event Log
  All files are up-to-date (2 minutes ago) 58841 CRLF+ UTF-8+
  1024 PM 6/24/2016

```

Word Count:

Input file:

```

// SparkWordCount - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\Spark WordCount] - input - IntelliJ IDEA 15.0.3
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
SparkWordCount input
Project Structure Run Terminal Event Log
  Project SparkWordCount [root] C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\Spark WordCount
    src
      project [root-build] sources root
        main
          java
          resources
        scala
          SparkWordCount.scala
          scala-2.11
    test
    target
    build.sbt
    input
    External Libraries
  Run SparkWordCount
    16/06/24 22:19:25 INFO MemoryStore: MemoryStore cleared
    16/06/24 22:19:25 INFO BlockManager: BlockManager stopped
    16/06/24 22:19:25 INFO BlockManagerMaster: BlockManagerMaster stopped
    16/06/24 22:19:25 INFO SparkContext: Successfully stopped SparkContext
    16/06/24 22:19:25 INFO ShutdownHookManager: Shutdown hook called
    16/06/24 22:19:25 INFO ShutdownHookManager: Deleting directory C:\Users\ndar\AppData\Local\Temp\spark-9f33d328-818b-49ab-baad-3253551633bf
    16/06/24 22:19:25 INFO RemoteActorRefProvider$remotingTerminator: Shutting down remote daemon.
    Process finished with exit code 0
  Favorites Terminal Messages Run Event Log
  SBT project import: [warn] Multiple dependencies with the same organization/name but different ver. (moments ago) * Scanning files to index
  Ask me anything 839 CRLF+ UTF-8+
  1024 PM 6/24/2016

```

Output file:

Project Report Increment 2

Spark WordCount - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\SparkWordCount] - [root] - ..\output\part-00000 - IntelliJ IDEA 15.0.3

```

    [RBPs, 0]
    (GMFB, 1)
    (gene, 1)
    (bone, 3)
    (carboxyl, 1)
    (BFR, 1)
    ((brain,, 1)
    (uridine, 2)
    (EDA, 1)
    (G2Mh1)
    (MRCI, 2)
    (acetolactate, 1)
    (MBP5, 2)
    (ligand, 5)
    (defensin,, 6)
    (convertase, 5)
    (R:interleukin, 3)
    (CRYZ, 1)
    (HMGCs, 4)
    (HCO, 2)
    (secretory, 1)
    (B:B, 1)
    (KNG, 2)
    (PRLR, 1)
    (:GLI, 3)
  
```

Run: SparkWordCount
Process finished with exit code 0

Terminal: 16/06/24 22:19:25 INFO MemoryStore: MemoryStore cleared
16/06/24 22:19:25 INFO BlockManager: BlockManager stopped
16/06/24 22:19:25 INFO BlockManagerMaster: BlockManagerMaster stopped
16/06/24 22:19:25 INFO OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/06/24 22:19:25 INFO SparkContext: Successfully stopped SparkContext
16/06/24 22:19:25 INFO ShutdownHookManager: Shutdown hook called
16/06/24 22:19:25 INFO ShutdownHookManager: Deleting directory C:\Users\ndar\AppData\Local\Temp\spark-9f33d328-018b-49ab-baad-3253551633bf
16/06/24 22:19:25 INFO RemoteActorRefProvider\$RemotingTerminator: Shutting down remote daemon.
Process finished with exit code 0

Event Log: 2 Event Log

SBT project import: [warn] Multiple dependencies with the same organization/name but different versions. To avoid conflict, pick one version: // [warn] * org.scala-lang:scala-compiler:2.11.0,2.11.8 // [warn] * org.scala-lang:mo_ (a minute ago) 1:1 LF+ UTF-8 6/24/2016 10:24 PM

Spark WordCount - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\SparkWordCount] - [root] - ..\output\part-00001 - IntelliJ IDEA 15.0.3

```

    (A:T, 1)
    (translocator,, 3)
    (GCN, 2)
    (CCN:cyclin, 1)
    (BFS, 1)
    (iphosphorylase, 4)
    (ABHD, 2)
    (DTNA, 1)
    ((gp, 1)
    (PIGA, 1)
    (iproteoglycan, 3)
    (angiotensin, 3)
    (end, 2)
    (Spelmyer, 1)
    (TNPO, 2)
    (KRR, 2)
    (acetylgalactosaminidase,, 1)
    (KIFAP, 2)
    (PPARG:peroxisome, 1)
    (CFL, 2)
    (deacetylase(N, 2)
    (HP:rhaptoglobin, 1)
    (FXRD, 4)
    (TAX, 4)
    (less, 2)
  
```

Run: SparkWordCount
Process finished with exit code 0

Terminal: 16/06/24 22:19:25 INFO MemoryStore: MemoryStore cleared
16/06/24 22:19:25 INFO BlockManager: BlockManager stopped
16/06/24 22:19:25 INFO BlockManagerMaster: BlockManagerMaster stopped
16/06/24 22:19:25 INFO OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/06/24 22:19:25 INFO SparkContext: Successfully stopped SparkContext
16/06/24 22:19:25 INFO ShutdownHookManager: Shutdown hook called
16/06/24 22:19:25 INFO ShutdownHookManager: Deleting directory C:\Users\ndar\AppData\Local\Temp\spark-9f33d328-018b-49ab-baad-3253551633bf
16/06/24 22:19:25 INFO RemoteActorRefProvider\$RemotingTerminator: Shutting down remote daemon.
Process finished with exit code 0

Event Log: 2 Event Log

SBT project import: [warn] Multiple dependencies with the same organization/name but different versions. To avoid conflict, pick one version: // [warn] * org.scala-lang:scala-compiler:2.11.0,2.11.8 // [warn] * org.scala-lang:mo_ (a minute ago) 9:7 LF+ UTF-8 6/24/2016 10:24 PM

Process execution:

Spark WordCount - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\SparkWordCount] - [root] - ...output\part-00001 - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

Spark WordCount > output > part-00001

Project Structure

input x SparkWordCount.scala x _SUCCESS x part-00000 x part-00001 x

SparkWordCount [root] (.idea, output, _SUCCESS, _part-00000, _part-00001, _SUCCESS, part-00000, part-00001)

project [root-build] (sources root)

src

main

java

resources

scala

SparkWordCount

scala-2.11

test

target

build.sbt

input

External Libraries

Run SparkWordCount

Process finished with exit code 0

Terminal Messages Run TODO Event Log

SBT project import: [warn] Multiple dependencies with the same organization/name but different versions. To avoid conflict, pick one version: // [warn] * org.scala-lang:scala-compiler:2.11.0,2.11.8 // [warn] * org.scala-lang:scala-compiler:2.11.0,2.11.8 (a minute ago) 33 chars 1481 LF+ UTF-8 's' 1025 PM 6/24/2016

Ask me anything

Sentiment Analysis:

SentimentAnalysisCoreNLP - [C:\Users\ndar\Desktop\Acads\KDM\Tutorial\Code\SentimentAnalysisCoreNLP] - [sentimentanalysiscorenlp] - ...src\main\java\TweetWithSentiment.java - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SentimentAnalysisCoreNLP > src > main > java > TweetWithSentiment

Project Structure

sentimentanalysiscorenlp [sentimentanalysiscorenlp-build]

src

main

java

SentimentAnalyzer

TweetWithSentiment

resources

scala

scala-2.11

test

target

build.sbt

External Libraries

Run SentimentAnalyzer

C:\Program Files\Java\jdk1.8.0_65\bin\java" ...

Adding annotator tokenize

Adding annotator split

Adding annotator parse

Loading parser from serialized file edu/stanford/nlp/models/lexparser/englishPCFG.ser.gz ... done [2.2 sec].

Adding annotator sentiment

TweetWithSentiment [line=NAME DESCRIPTION ALL TA+ Norel ALL TA+ (NL) Norel ALL Norel ALL TA Norel ALL TA Norel ALL R (ML) Rel

ALPPL ALPPL :alkaline phosphatase, placental like

"RPLP_RPLP :ribosomal protein, large, P

"GFER_GFER:growth factor, augmenter of liver regeneration (ERV homolog, S cerevisiae)

"IGHM_IGHM:immunoglobulin heavy constant mu

"SART_SART :squamous cell carcinoma antigen recognised by T cells

"RPLP_RPLP :ribosomal protein, large, P , cssClass=sentiment : very negative]

Process finished with exit code 0

Terminal Messages Run TODO Event Log

SBT project import: [warn] Multiple dependencies with the same organization/name but different versions. To avoid conflict, pick one version: // [warn] * org.scala-lang:scala-compiler:2.11.0,2.11.8 // [warn] * org.scala-lang:scala-compiler:2.11.0,2.11.8 (2 minutes ago) 565 CRLF+ UTF-8 's' 1025 PM 6/24/2016

Ask me anything

8 PROJECT MANAGEMENT

8.1 CONTRIBUTION:

Overall – 100%

Venkata Vamsi Krishna Bhuvanam – 25%

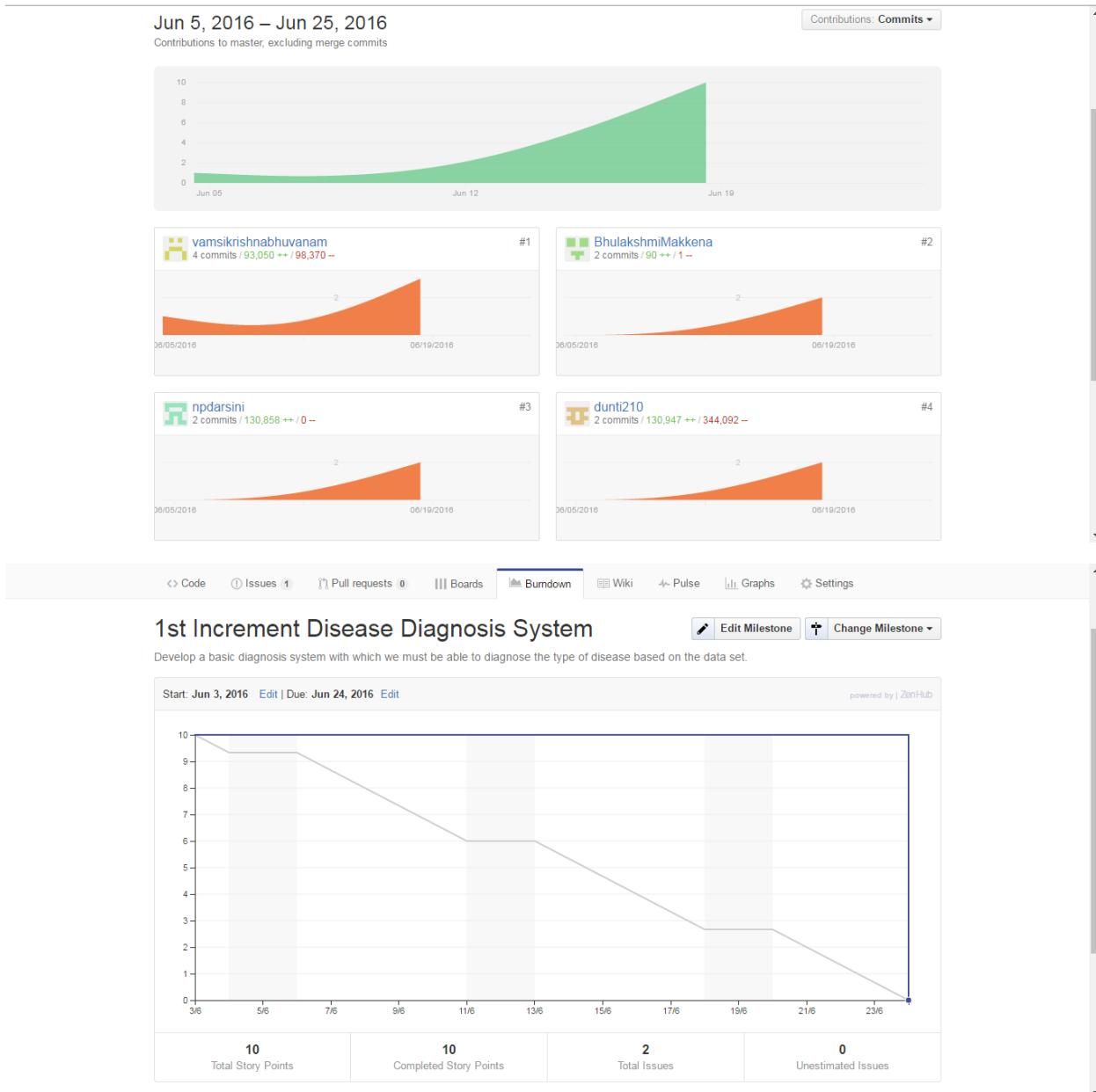
Priyadarsini Nidadavolu – 25%

Bhulakshmi Makkena – 25%

Tej Kumar Yentrapragada – 25%

8.2 ZENHUB AND GITHUB:

8.2.1 First Increment:



Project Report Increment 2

No description or website provided. — Edit

Branch: master ▾ New pull request

Create new file Upload files Find file Clone or download ▾

dunti210 Commit Latest commit 2c50b8d 10 minutes ago

1st Increment/Source/Spark-TF-IDFChanged Commit 4 minutes ago

1stIncrement Commit 4 minutes ago

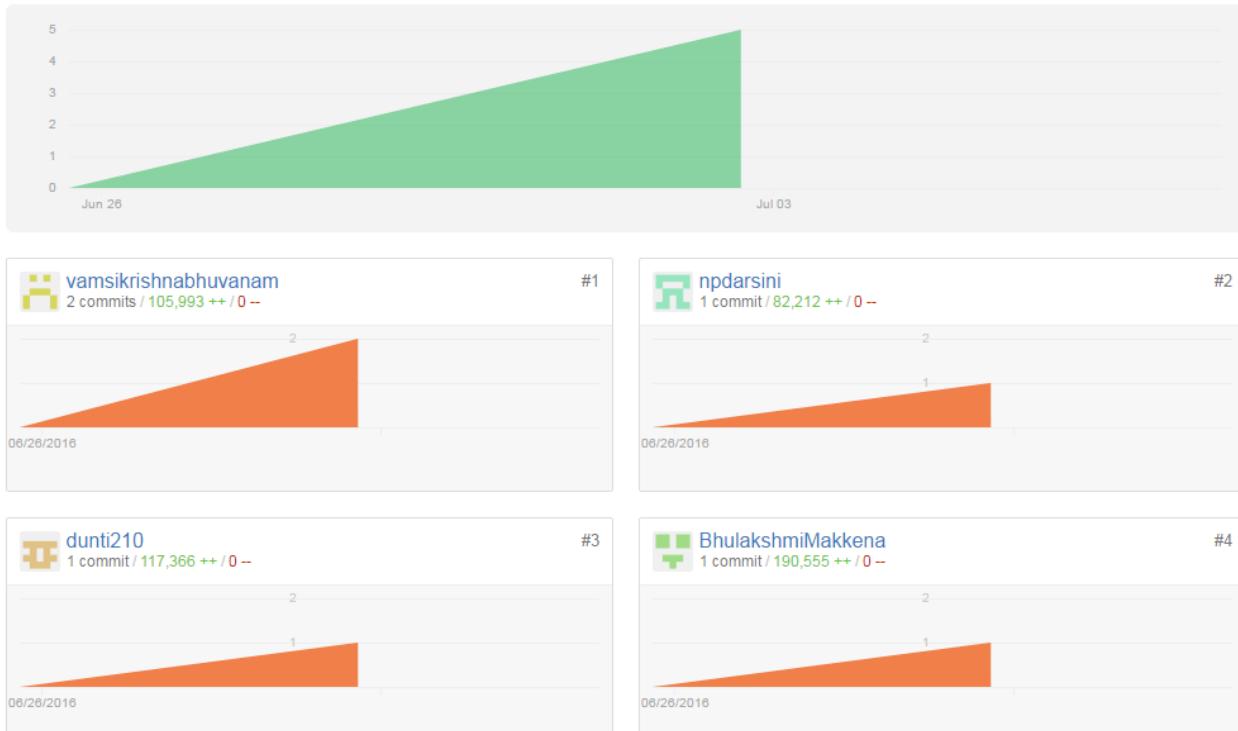
README.md Initial commit 17 days ago

README.md

KDM-Summer-2016-ProjectDraft

8.2.2 SECOND INCREMENT:

Contributions to master, excluding merge commits



This screenshot shows the GitHub repository page for 'Summer2016-KDM / KDM-Summer-2016-Strikerz'. The page includes standard GitHub navigation like 'Unwatch', 'Star', 'Fork', and 'Settings'. Below the navigation, there are tabs for 'Code', 'Issues (0)', 'Pull requests (0)', 'Boards', 'Burndown', 'Wiki', 'Pulse', 'Graphs', and 'Settings'. The main content area displays repository statistics: 5 commits, 1 branch, 0 releases, and 1 contributor. It also shows the latest commit by npdarsini and a list of files including '2nd Increment/Source', 'README.md', and 'README.md' (a duplicate entry).

Created a new Repository because of contribution issues in Earlier Repo. This repository consists from 2nd Increment. — Edit

This screenshot shows the GitHub repository page for 'KDM-Summer-2016-Strikerz'. The repository has 5 commits, 1 branch, 0 releases, and 1 contributor. The latest commit was made 6 minutes ago by npdarsini. The repository contains files such as '2nd Increment/Source', 'README.md', and 'README.md' (a duplicate entry). The README file content is identical to the one in the previous repository, stating: 'Created a new Repository because of contribution issues in Earlier Repo. This repository consists from 2nd Increment.'

8.3 CONCERNS/ISSUES:

NA

8.4 FUTURE WORK:

Concentrating to continue to work on Drug Dataset collection and continue to build a recommendation system which will recommend drug to the user.