* Parsing the Crawled News
* Indexing the root words & stop words
* Extracting the keywords
* Categorizing the news
* Extracting the location

**Parsing the Crawled News:** We have used Jsoup 1.7.1 API to parse the whole crawled data. We got more than 500 MB crawled news Data from Pipilika. Every news was stored with syntax of XML and they were well formatted in packet by packet. Here is the example of a news packet.

|  |
| --- |
| <Index>  <byteInfo>1 11315933 11315980 11315980 11342687</byteInfo>  <indexed>true</indexed>  <TITLE>নরসিংদীতে ছাত্রলীগ নেতা খুন</TITLE>  <CONTENT>নরসিংদী জেলা ছাত্রলীগের এক নেতাকে কুপিয়ে হত্যা করেছে দুর্বত্তরা। মঙ্গলবার বিকালে শিবপুর উপজেলার পুটিয়া  বাজার এলাকায় এ হামলায় নিহত শাহিন কাদির মাহিন ওরফে মাইনুদ্দিন (৩০) জেলা ছাত্রলীগের সাংগঠনিক সম্পাদক  বলে পুলিশ জানিয়েছে। সদর থানার ওসি মো. আসাদুজ্জামান জানান, দুপুর ২টার দিকে মাহিন পুটিয়ার বাজারে নিজ  ব্যবসা প্রতিষ্ঠানে যান। সেখান থেকে বিকালে ঘোড়াদিয়া এলাকায় বাড়ি ফেরার পথে কয়েকজন সন্ত্রাসী তাকে কুপিয়ে  গুরুত্বর আহত করে। নরসিংদী জেলা সদর হাসপাতালে নেয়ার পর কর্তব্যরত চিকিৎসক তাকে মৃত ঘোষণা করেন।  মাহিনের মাথা ও গায়ে একাধিক কোপের চিহ্ন রয়েছে। তার মরদেহ এখন জেলা সদর হাসপাতালে আছে বলে জানান  ওসি। কে বা কারা এ হামলা করেছে তা জানা না গেলেও নিহতের পরিবারের দাবি, রাজনৈতিক কারণে তাকে হত্যা করা  হয়েছে।  </CONTENT>  <CATEGORY>CRIME POLITICS DESH</CATEGORY>  <CITY>ঢাকা</CITY>  <DOMAIN>http://www.kalerkantho.com</DOMAIN>  <DATE>201404250000</DATE>  <URL>http://www.kalerkantho.com/feature/ronger-mela/2014/04/24/76173</URL>  <TYPE>news</TYPE>  </Index> |

From this news packet we need title, content and date. But we parsed all these news features for our performance measurement. Here we parsed title, content, category, city, domain, date, url and type. Here I just became careful about the date and we parsed date feature manually. Here is the source code for parsing news packet.

|  |
| --- |
| Void ParseDocument(String data)  {  Document doc = Jsoup.parse(data, "", Parser.xmlParser());  Elements el = doc.select("Index");  String title, content, category, city, domain, date, url, type;  for (Element e : el) {  title = (e.select("TITLE").text());  content = (e.select("CONTENT").text());  category = (e.select("CATEGORY").text());  city = (e.select("CITY").text());  domain = (e.select("DOMAIN").text());  date = (e.select("DATE").text());  url = (e.select("URL").text());  type = (e.select("TYPE").text());  date = PurifyDateString(date);  //Do whatever I want  }  }  private String PurifyDateString(String date) {  if (date.contains("-"))  {  String[] str = date.split("[ -]+");  dd = str[2] + "-" + str[1] + "-" + str[0];  }  else  {  dd += date.substring(0, 4); dd += "-";  dd += date.substring(4, 6); dd += "-";  dd += date.substring(6, 8);  }  return dd;  } |

**Indexing Root words and Stop word:** Bengali has as many as 1,00,000 unique words, of which 50,000 are considered ‘Totsomo’ (তৎসম- direct re-borrowings from Sanskrit), 21,100 are ‘Todbhobo’ (তদ্ভব- native words) and the rest being ‘Bideshi’ (বিদেশী foreign borrowings) and ‘Deshi’ (দেশী- Austroasiatic borrowings). All Bengali words are categorized into 5 parts of speech: Noun, Adjective, Pronoun, Conjunction and Verb. Nouns and Pronouns are inflected for case, including nominative, objective, genitive (possessive) and locative. The case marking pattern for each noun being inflected depends on the noun’s degree of animacy. When a definite article such as – টা-ta(Singular) or গুলা-gula (plural) is added, as in the tables below, nouns are also inflected for number.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Singular noun inflection** | | | |  | **Animate** | **Inanimate** | | **Nominative** | ছাত্রটা chhatrô-ţa the student | জুতাটা juta-ţa the shoe | | **Objective** | ছাত্রটা**কে** chhatrô-ţa-**ke** the student | জুতাটা juta-ţa the shoe | | **Genitive** | ছাত্রটা**র** chhatrô-ţa-**r** the student's | জুতাটা**র** juta-ţa-**r** the shoe's | | **Locative** | - | জুতাটা**য়** juta-ţa-**y** on/in the shoe | | |  |  |  | | --- | --- | --- | | **Plural noun inflection** | | | |  | **Animate** | **Inanimate** | | **Nominative** | ছাত্র**রা** chhatrô-**ra** the students | জুতাগুলা/জুতোগুলো juta-gula/juto-gulo the shoes | | **Objective** | ছাত্র**দের(কে)** chhatrô-**der(ke)** the students | জুতাগুলা/জুতোগুলো juta-gula/juto-gulo the shoes | | **Genitive** | ছাত্র**দের** chhatrô-**der** the students' | জুতাগুলা/জুতোগুলো**র** juta-gula/juto-gulo-**r** the shoes' | | **Locative** | - | জুতাগুলা/জুতোগুলো**তে** juta-gula/juto-gulo-**te** on/in the shoes | |

A document has a list of words containing useless words and useful words. For example we may look over a sample Bengali Article:

|  |
| --- |
| Sample Bengali Article: ঢাকার বনানী এলাকায় **এক** বিকাশ প্রতিনিধিকে কুপিয়ে ৭ লাখ টাকা ছিনতাই করেছে দুর্বৃত্তরা। রোববার দুপুর দেড়টায় রাজধানীর সবুজবাগ থানার বাসাবো ওয়াসা রোডে জিন ইন্টারন্যাশনাল নামের বিকাশ এজেন্সির বিক্রয় প্রতিনিধি এনামুল হককে (৪০) কুপিয়ে সাত লাখ টাকা ছিনতাই করে। তিনি ঢাকা মেডিকেল কলেজ হাসপাতালে (ঢামেক) চিকিৎসাধীন রয়েছেন। এনামুল হক জানান, ওই স্থানে ৫/৬ দুর্বৃত্ত তার গতিরোধ করে ধারালো অস্ত্র দিয়ে কুপিয়ে হাতে থাকা টাকার ব্যাগটি নিয়ে যায়। ব্যাগে সাত থেকে আট লাখ টাকা ছিল বলে তিনি জানান। ঢামেকের কর্তব্যরত চিকিৎসক জানান, এনামুল হকের হাতে, ‍বুকে ও পিঠে ধারালো অস্ত্রের আঘাত আছে। জিন ইন্টারন্যাশনালের সুপারভাইজার তানভির নেওয়াজ খান জানান, খবর পেয়ে প্রথমে এনামুল হককে বাসাবো জেনারেল হাসপাতালে নিয়ে যাওয়া হয়। পরে উন্নত চিকিৎসার জন্য তাকে ঢামেকে স্থানান্তর করা হয়। সবুজবাগ থানার ভারপ্রাপ্ত কর্মকর্তা (ওসি) বাবুল মিয়া এ প্রসঙ্গে জানান, মারামারির ঘটনা শুনেছি। ছিনতাই কীনা খতিয়ে দেখা হচ্ছে। |

Here we have found out a set of stop words: {“এক”, “নামের”, “করে”, “তিনি”, “রয়েছেন”, “জানান”, “তার”, “করে”, “দিয়ে”, “থাকা”, “নিয়ে”, “যায়”, “ছিল”, “বলে”, “আছে”, “প্রথমে”, “যাওয়া”, “হয়”, “পরে”, “করা”, “শুনেছি”, “কিনা”, “দেখা”, “হচ্ছে”}. And the rest words are considered as main word (Keyword). We have studied more than 500 news articles and stored more than 1000 stop words in a text file.

Storing root word is not so easier like stop word. There are more than 40,000 words are used in bangladeshi news articles. In order to computer root word from a sample word we have to study about bangla grammartical rules (পদ, প্রকৃতি, প্রত্যয়, উপসর্গ, কারক, বিভক্তি, সন্ধি-বিচ্ছেদ). But root word finding was not our concern. So we stored more than 4,500 words and corresponding root word in a text file. The root word map is shown below:

|  |
| --- |
| পুলিশে পুলিশ  পুলিশকে পুলিশ  পুলিশের পুলিশ  পুলিশও পুলিশ  পুলিশদের পুলিশ  পুলিশরা পুলিশ  পুলিশদেরকে পুলিশ  খেলায় খেলা  খেলার খেলা  খেলতে খেলা  খেলাকে খেলা  খেলাগুলো খেলা |

Extracting Keywords: **We have studied about morethan 500 news articles and observed that every Pronoun(সর্বনাম) and Conjunction(অব্যয়) are stop word. Pronouns and Conjunctions are not used to define a news category. Pronouns and Conjunction are added to stop word list. We can also add several verb(ক্রিয়া) and Adjective(বিশেষন) words to stop word list.** Actually we need all the Nouns(বিশেষ্য), several verbs(ক্রিয়া) and adjectives**(বিশেষন)**.

We picked keyword from about 200 news articles. First of all, we counted term frequency for every term. The term which is not a stop word and exist more than 5 times we picked it as a keyword. Here is the sample term frequencies of our news data.

|  |
| --- |
| পুলিশ -->crime = 200 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 9 Total: 209  দল -->crime = 24 -->sports = 61 -->entertainment = 1 -->technology = 1 -->others = 31 Total: 118  ডাকাত -->crime = 72 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 72  মামলা -->crime = 55 -->sports = 0 -->entertainment = 2 -->technology = 2 -->others = 2 Total: 61  শিশু -->crime = 32 -->sports = 0 -->entertainment = 2 -->technology = 0 -->others = 25 Total: 59  খুন -->crime = 55 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 55  অভিনয় -->crime = 1 -->sports = 0 -->entertainment = 50 -->technology = 0 -->others = 0 Total: 51  হাসপাতাল -->crime = 44 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 6 Total: 50  ছিনতাই -->crime = 50 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 50  কর্মকর্তা -->crime = 34 -->sports = 2 -->entertainment = 0 -->technology = 5 -->others = 9 Total: 50  হত্যা -->crime = 49 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 49  অভিযোগ -->crime = 40 -->sports = 0 -->entertainment = 1 -->technology = 0 -->others = 5 Total: 46  ম্যাচ -->crime = 1 -->sports = 44 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 45  আহত -->crime = 41 -->sports = 0 -->entertainment = 2 -->technology = 0 -->others = 2 Total: 45  বিশ্বকাপ -->crime = 0 -->sports = 40 -->entertainment = 1 -->technology = 0 -->others = 0 Total: 41  নিহত -->crime = 37 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 3 Total: 40  গোল -->crime = 0 -->sports = 35 -->entertainment = 0 -->technology = 0 -->others = 2 Total: 37  লাশ -->crime = 34 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 1 Total: 35  ধর্ষণ -->crime = 30 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 2 Total: 32  মেয়ে -->crime = 20 -->sports = 2 -->entertainment = 9 -->technology = 0 -->others = 0 Total: 31  স্ত্রী -->crime = 24 -->sports = 2 -->entertainment = 2 -->technology = 0 -->others = 2 Total: 30  খেলা -->crime = 3 -->sports = 23 -->entertainment = 2 -->technology = 0 -->others = 2 Total: 30  আসামি -->crime = 30 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 30  রিয়াল -->crime = 0 -->sports = 27 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 27  কোচ -->crime = 0 -->sports = 27 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 27  ফাইনাল -->crime = 0 -->sports = 25 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 25  গুলি -->crime = 24 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 24  অভিনেত্রী -->crime = 0 -->sports = 0 -->entertainment = 22 -->technology = 0 -->others = 2 Total: 24  মাঠ -->crime = 4 -->sports = 16 -->entertainment = 1 -->technology = 1 -->others = 0 Total: 22  পয়েন্ট -->crime = 2 -->sports = 14 -->entertainment = 1 -->technology = 0 -->others = 5 Total: 22  দুর্বৃত্ত -->crime = 22 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 22  গান -->crime = 0 -->sports = 0 -->entertainment = 18 -->technology = 0 -->others = 4 Total: 22  অস্ত্র -->crime = 21 -->sports = 1 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 22  হামলা -->crime = 18 -->sports = 0 -->entertainment = 0 -->technology = 1 -->others = 2 Total: 21  মাইক্রোসফট -->crime = 0 -->sports = 0 -->entertainment = 0 -->technology = 21 -->others = 0 Total: 21  অভিনেতা -->crime = 0 -->sports = 0 -->entertainment = 20 -->technology = 0 -->others = 0 Total: 20  জয় -->crime = 1 -->sports = 13 -->entertainment = 1 -->technology = 0 -->others = 3 Total: 18  খেলোয়াড় -->crime = 0 -->sports = 18 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 18  মৃত্যু -->crime = 13 -->sports = 0 -->entertainment = 2 -->technology = 0 -->others = 2 Total: 17 |

Here we picked a term as a keyword which is more frequent in news articles.

$\displaystyle P(c\vert d) \propto P(c) \prod_{1 \leq \tcposindex \leq n_d} P(\tcword_\tcposindex\vert c)$**News Categorization Implementation:** We used Naïve Bayes text classification for news article categorization. The probability of a document d being in class c is computed as

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Here, P( c ) =

nd = Number of term in Document d

P(tk|c) =

\begin{displaymath}
\hat{P}(\tcword\vert c) = \frac{T_{c\tcword}}{\sum_{\tcword'
\in V} T_{c\tcword'}},
\end{displaymath}

Now we will compute probability for every category individually. Maximum value will be defined by specific category.\begin{displaymath}c_{map} =
\argmax_{\tcjclass \in \mathbb{C}} \hat{P}(\tcjc...
...posindex \leq n_d}
\hat{P}(\tcword_\tcposindex\vert\tcjclass).
\end{displaymath}

In this equation many conditional probabilities are multiplied, one for each position 1<=k<=nd. This can result in a floating point underflow. It is therefore better to perform the computation by adding logarithms of probabilities instead of multiplying probabilities. The class with the highest log probability score is still the most probable. So the equation will be changed to,

$\displaystyle c_{map} = \argmax_{\tcjclass \in \mathbb{C}} \ [ \log \hat{P}(\tc...
...{1 \leq \tcposindex \leq n_d}
\log \hat{P}(\tcword_\tcposindex\vert\tcjclass)].$

To eliminate zeros, we use add-one technique, which simply adds one to each count.

\begin{displaymath}
\hat{P}(\tcword\vert c) = \frac{T_{c\tcword}+1}{\sum_{\tcwor...
...frac{T_{c\tcword}+1}{(\sum_{\tcword' \in V} T_{c\tcword'})+B},
\end{displaymath}

Here is the source code of implementation of naive bayes clustering:

|  |
| --- |
| private String Naive\_Bayes\_Clustering(String document)  {    double Likelihood\_Probability, Posterior\_Probability;  double Prior\_Probability[] = new double[5];  Occurance O;  String words[] = document.split(" ");  for(int i=0;i<Prior\_Probability.length;i++)  {    Prior\_Probability[i]= Math.log10(documents\_kount[i]/ total\_documents);    for(String term:words)  {  if(rawData.useless.contains(term)) continue;  if(rawData.rootWord.containsKey(term))  term = rawData.rootWord.get(term);    if(preeData.TermFrequency.containsKey(term))  O = preeData.TermFrequency.get(term);  else  O = new Occurance();    Likelihood\_Probability = (O.kounter[i]+1)\*1.0 / (O.total + 1)\*1.0;  Prior\_Probability[i]+=Math.log10(Likelihood\_Probability);  }    }    int indx=0;  double argMax=-(1<<30);  for(int i=0;i<Prior\_Probability.length;i++)  {  if(Prior\_Probability[i]>argMax)  {  argMax = Prior\_Probability[i];  indx=i;  }  }    return getNameOf(indx);  }  private String getNameOf(int id)  {  if(id==0) return "crime";  if(id==1) return "sports";  if(id==2) return "entertainment";  if(id==3) return "technology";  return "others";    } |

Now, Data for parameter estimation example for Naïve Bayes Clustering:

|  |  |
| --- | --- |
|  |  |
| ১৪ বছর আগে চট্টগ্রামের বহদ্দারহাটে ছাত্রলীগের গাড়িতে হামলা চালিয়ে আট জনকে হত্যার য়ে মৃত্যুদণ্ডে দণ্ডিত চার আসামির সবাই আপিলের রায়ে খালাস পেয়েছেন। | Crime |
| রাজধানীতে দুর্ধর্ষ ডাকাতি । বাসার দুই দারোয়ানকে বেঁধে পার্কিং করা একটি পালসার ব্রান্ডের মোটরসাইকেল নিয়ে যায় ও একটি প্রাইভেটকারের (ঢাকা মেট্রো-গ-৩৫-৩৫৯৯) যন্ত্রাংশ খুলে নিয়ে যায়। | Crime |
| চুয়াডাঙ্গার দামুড়হুদায় গভীর রাতে বাসায় ঢুকে আমজাদ হোসেন (৪৫) নামের এক ব্যক্তিকে কুপিয়ে হত্যা করেছে সন্ত্রাসীরা। শুক্রবার রাতে উপজেলার জয়রামপুর গ্রামের চৌধুরী পাড়ায় এ ঘটনা ঘটে। নিহতের লাশ শনিবার সকাল ৭টার দিকে উদ্ধার করে ময়নাতদন্তের জন্য চুয়াডাঙ্গা সদর হাসপাতালের মর্গে পাঠিয়েছে পুলিশ। | Crime |
| টি-টোয়েন্টি বিশ্বকাপের চ্যাম্পিয়ন শ্রীলঙ্কা  এখনো পর্যন্ত টি-টোয়েন্টি বিশ্বকপের সবগুলো ম্যাচেই জয়লাভ করেছে ভারত।  আইসিসি টি২০ বিশ্বকাপের ফাইনালে ভারতকে ৬ উইকেটে হারিয়ে শিরোপা নিজেদের ঘরে নিয়েছে শ্রীলঙ্কা। | Sports |
| নতুন একটা সোনাক্ষি  বলিউডে প্রধান আলোচ্য এখন একটাই- ওজন কমিয়েছেন সোনাক্ষি সিনহা! এবং প্রেম করছেন শহিদ কাপুরের সঙ্গে। 'ডেইলি টেলিগ্রাফ' অবলম্বনে সোনাক্ষিকে নিয়ে লিখেছেন শাকিল ফারুক লোকের কৌতূহলের সীমারেখা থাকে না আসলে। নইলে ভারতের এমন নির্বাচনী ডামাডোলে কংগ্রেস-বিজেপি-আম আদমি পার্টিকে সরিয়ে, সোনাক্ষি সিনহার ওজন প্রসঙ্গ কি আর আলোচনার মুখ্য বিষয় হয়ে উঠতে পারে! সোনাক্ষি যত বিরক্তি নিয়েই বলুন, 'এ নিয়ে কথা বলতে আগ্রহী নই'- আলোচনা চলবেই। | Entertainment |
| অঅ-অ+ টিভি পর্দায় জাহিদ হাসানের নানা রূপ। পুরনো রূপ ভেঙে তিনি আবার নতুন রূপে হাজির হয়েছেন 'নজিরবিহীন নজির আলী' নাটকে। লিখেছেন মাহ্‌বুব হাসান জ্যোতি 'আমি ভাই ভিন্ন ধরনের চরিত্রের কাঙাল'- আলাপচারিতার শুরুতেই বললেন জাহিদ হাসান। আরমান ভাইয়ের চরিত্র ছাপিয়ে তিনি এখন নজর আলী হয়ে উঠছেন। | Entertainment |
| বিশ্বকাপের সঙ্গে ফ্রান্সের কোচের ভাবনায় ইউরো  ইউরোপের সবচেয়ে বড় ফুটবল আসরের আয়োজক ফ্রান্স। তাই এখন থেকেই ভবিষ্যত সাফল্যের কথা ভাবতে হচ্ছে ফরাসি এই কোচকে। | Sports |
| অর্থনৈতিক উন্নতি ও নৈতিক দায়  ভোরবেলা যাচ্ছিলাম কমলাপুর স্টেশনে। রাস্তায় খুব কম যানবাহন। তবু এক সিগন্যালে একটু থামতে হলো বাঁ দিক থেকে কয়েকটি গাড়ি ক্রসিং পার হয়ে ডান দিকে আসায়। একটি পিকআপ পাশ ঘেঁষে থামতে বাধ্য হলো। অত ভোরে কোনো বোকা চালকও লালবাতি মানেন না। পিকআপে দুটো চটের বস্তা। একটি বস্তার মুখের দিকে সামান্য ফাঁক দিয়ে দেখা গেল একটি মরা মুরগির পা ও পাখনা। চল্লিশ-পঞ্চাশ বছর আগে হলে মনে করতাম মরা মুরগি কুড়িয়ে কেউ ডাস্টবিনে ফেলে দিতে নিয়ে যাচ্ছে। এখন মনে পড়ল অন্য কথা। আমার কর্তব্য ছিল গাড়িটিকে আটকে পুলিশকে খবর দেওয়া। তা না করতে পারায় গ্লানি ও অপরাধ বোধ করি। | Others |
| গ্যালাক্সি এস৫ মিনি আসছে  গ্যালাক্সি এস৫ স্মার্টফোনটির একটি মিনি বা ছোট সংস্করণ আসছে। গ্যালাক্সি এস৫ এর এ সংস্করণটি হবে পানি-রোধী। স্যামসাং নিউজিল্যান্ডের অফিশিয়াল ওয়েবসাইটে এ তথ্য জানানো হয়েছে। অবশ্য মিনি সংস্করণটির তথ্য এখনও আনুষ্ঠানিকভাবে ঘোষণা করেনি দক্ষিণ কোরিয়ার প্রতিষ্ঠানটি। | Technology |

Now here is the data structure for calculating Term Frequency from learned data.

|  |
| --- |
| সোনাক্ষি -->crime = 0 -->sports = 0 -->entertainment = 4 -->technology = 0 -->others = 0 Total: 4  বিশ্বকাপ -->crime = 0 -->sports = 4 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 4  ভারত -->crime = 0 -->sports = 2 -->entertainment = 1 -->technology = 0 -->others = 0 Total: 3  গ্যালাক্সি -->crime = 0 -->sports = 0 -->entertainment = 0 -->technology = 3 -->others = 0 Total: 3  হত্যা -->crime = 2 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 2  শ্রীলঙ্কা -->crime = 0 -->sports = 2 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 2  রূপ -->crime = 0 -->sports = 0 -->entertainment = 2 -->technology = 0 -->others = 0 Total: 2  রাত -->crime = 2 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 2  মরা -->crime = 0 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 2 Total: 2  মন -->crime = 0 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 2 Total: 2  ফ্রান্স -->crime = 0 -->sports = 2 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 2  পুলিশ -->crime = 1 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 1 Total: 2  তথ্য -->crime = 0 -->sports = 0 -->entertainment = 0 -->technology = 2 -->others = 0 Total: 2  টোয়েন্টি -->crime = 0 -->sports = 2 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 2  জাহিদ -->crime = 0 -->sports = 0 -->entertainment = 2 -->technology = 0 -->others = 0 Total: 2  চরিত্র -->crime = 0 -->sports = 0 -->entertainment = 2 -->technology = 0 -->others = 0 Total: 2  গাড়ি -->crime = 1 -->sports = 0 -->entertainment = 0 -->technology = 0 -->others = 1 Total: 2  কোচ -->crime = 0 -->sports = 2 -->entertainment = 0 -->technology = 0 -->others = 0 Total: 2 |

Here is a new document:

|  |
| --- |
| ডাকাতি মামলার আসামিকে কুপিয়ে হত্যা  ঢাকার কেরানীগঞ্জে বাবুল ওরফে হক বাবুল (৩২) নামের এক ব্যক্তিকে কুপিয়ে হত্যা করেছে দুর্বৃত্তরা। |

Here is the calculated probabilities using this algorithm:

|  |
| --- |
| crime: -0.47712125471966244  sports: -3.112605001534575  entertainment: -3.112605001534575  technology: -3.4136349971985562  others: -3.4136349971985562  Category Calculated: crime (As it is the Maximum Probability) |