CIS 635 - Knowledge Discovery & Data Mining

Sequence data and modeling introduction

Sequence data

- NLP
  - Machine Translation (MT)
  - Question Answering
  - Document Classification
  - Sentiment Classification
  - Document summarization

- DNA Sequencing
  - DNA sequencing
  - DNA classification

Sequence data

- NLP
  - Machine Translation (MT)
  - Question Answering
  - **Document Classification**
  - Sentiment Classification
  - Document summarization

- DNA Sequencing
  - DNA sequencing
  - **DNA classification**

Documents

The healthcare industry is a little different because not much marketing needs to be done for you to earn more patients. After all, health is wealth and people want to keep their health in check.

However, ***this doesn't mean you can engage with your patients any less***. In fact, helping them with the little things is what's important, such as sending them a simple message to remind them about their upcoming appointment.

But, while sending messages to your customers can help improve engagement and resolve queries faster, having to manually send individual messages can be extremely time-consuming.

This is why you need message templates to help speed up the process and cut down on manual work. Furthermore, you don't have to reinvent the wheel for every conversation with a patient!

Let's take a look at the best medical message templates so you can send more messages in less time. 👇

12 best medical message templates

Message templates for Appointment Reminders

No-shows are inevitable. While they are challenging to eliminate completely, you can definitely reduce no-show rates by sending them a simple reminder text about their upcoming appointments.

Essay By SpeedyPaper

Source: https://speedypaper.com/essays/challenges-in-sports

Challenges in Sports, Essay Sample for Students Essay

For many people, athletics has served as a powerful positive force in their lives. Parents who encourage their kids to take up sports, for example, do it in the hope that the sport they engage in will be beneficial to them in the end. Nevertheless, there are significant challenges that come along with sports. The most common include criminal activity among players and violence among spectators and fans in general. Both of these issues have continuously posed great threats to the dignity that sport was meant to have. There is a need to go through all the essentials of our main form of leisure. There is a need to reverse it to the highly respected activity it was. This paper details how violence and criminal activity have played degrading roles as regards to sports.

Since the early 1900s, professional sports have become popular as the perfect choice of entertainment for many Americans. While employment and production in the U.S. manufacturing and service industries fluctuated during the business cycles that occurred between 1900 and 1950, American professional sports leagues tended to flourish. The activity has become extremely deep-rooted in more countries and regions such that many sports lovers proudly idolize ballplayers and track athletes for their persistent hustle and unique athleticism. Some fans even go as far as trying to emulate the sporting moves of their favorite athletes such as the famous home run swing of San Francisco Giants outfielder Barry Bonds, the thunderous dunk of Lakers center Shaquille ONeal, and the pinpoint passes of Green Bay Packers quarterback Brett Favre. Others collect sports cards and exchange them before hosting tallage parties before and after home games. They manage teams in fantasy sports, participate in local campaigns, and convince voters to adopt stadium referendums, and watch sports channels at sports bars. Consequently, professional team sports become a frequent and significant leisure activity for a vast segment of the U.S. population.

Throughout the last few years, the sports coverage has mainly constituted crimes committed by athletes or disorderly fans. This type of media coverage has ensured that sports have dominated many conversations across the country and the rest of the world. It is not a surprise that polls have also indicated that during the same time, crime has been the most popular subject of the media. The once irrelevant and not so fan-filled sports page have now turned into a crime reports page. It is now much easier to turn the paper and begin with the sports page because its stories have become juicier with time. And since it is the duty of the press to express themselves freely, one need only open the newspaper to read about the Barry Bonds steroids debate, Michael Vicks dogfighting conviction or the fan and player fight during the Detroit Pistons and Indiana Pacers basketball game. Athletics have always held a special place in many communities all across the country. As a result, it is now a proven fact that influence on athletic behavior begins in high school. Towns continuously support athletic programs through fundraisers, rallies, and event attendance. This is clearly evident since currently, nothing is much more important to the American people than their weekend soccer events, Friday night football games in high school, or the many baseball tournaments.

The community has long been speaking of teamwork, competition, and fitness to their sons and daughters but this positivity has only brought with it darker sides to the sporting world. Are the stories in the paper an anomaly or is there something inherent in some sports or team that encourages criminal behavior? Despite the overemphasis on high school sports, athletes tend to maintain their special place in society even after joining college. They are supposed to develop some

medicine and cultivate meaningful relationships with patients and their families.

Sequence data

- **Data/Feature encoding**
  - One-Hot Encoding
  - Label Encoding

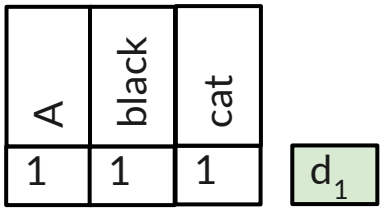
What are the challenges?

Sequence data

- **Data/Feature encoding**
  - One-Hot Encoding
  - Label Encoding

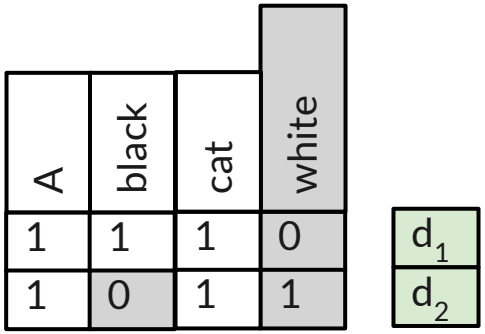
- **NLP/DNA sequencing**
  - Tf-idf
  - **CountVectorizer**

CountVectorizer – general idea



“A black cat”

CountVectorizer – general idea



“A black cat”

“A white cat”

CountVectorizer – general idea

A	black	cat	white	is	as	beautiful	
1	1	1	0	0	0	0	d_1
1	0	1	1	0	0	0	d_2
2	1	2	1	2	2	1	d_3

“A black cat”

“A white cat”

“A black cat is as beautiful as a white cat is”

CountVectorizer – general idea

A	black	cat	white	is	as	beautiful	
1	1	1	0	0	0	0	d_1
1	0	1	1	0	0	0	d_2
2	1	2	1	2	2	1	d_3

Corpus
“A black cat”
“A white cat”
“A black cat is as beautiful as a white cat is”

CountVectorizer – general idea

A	black	cat	white	is	as	beautiful	
1	1	1	0	0	0	0	d_1
1	0	1	1	0	0	0	d_2
2	1	2	1	2	2	1	d_3

Corpus

"A black cat"

"A white cat"

"A black cat is as beautiful as a white cat is"

dictionary: {
"a", "is", "as",
"cat", "black",
"white", "beautiful"
}

CountVectorizer – general idea

A	black	cat	white	is	as	beautiful	
1	1	1	0	0	0	0	d_1
1	0	1	1	0	0	0	d_2
2	1	2	1	2	2	1	d_3

Corpus

"A black cat"

"A white cat"

"A black cat is as beautiful as a white cat is"

Unigram: *["a", "is", "as", "cat", "black", "white", "beautiful"]*

CountVectorizer – general idea

A	black	cat	white	is	as	beautiful	
1	1	1	0	0	0	0	d_1
1	0	1	1	0	0	0	d_2
2	1	2	1	2	2	1	d_3

Corpus

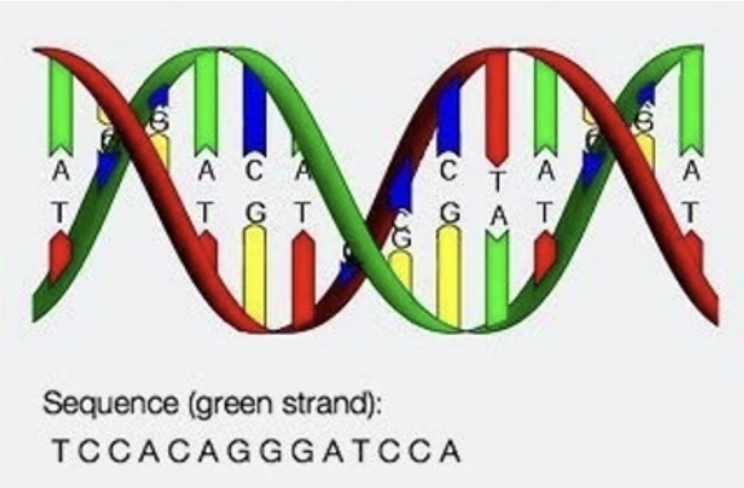
"A black cat"

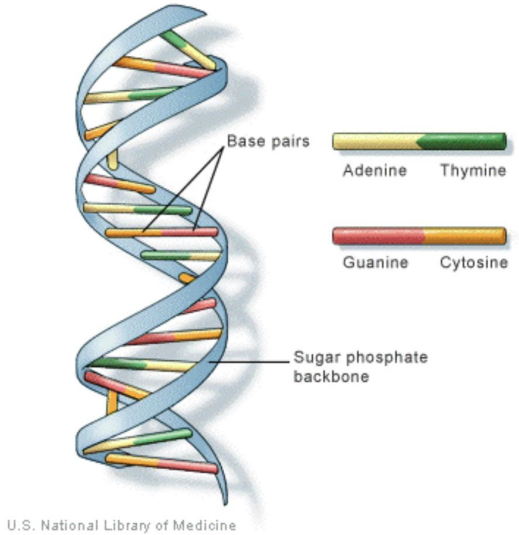
"A white cat"

"A black cat is as beautiful as a white cat is"

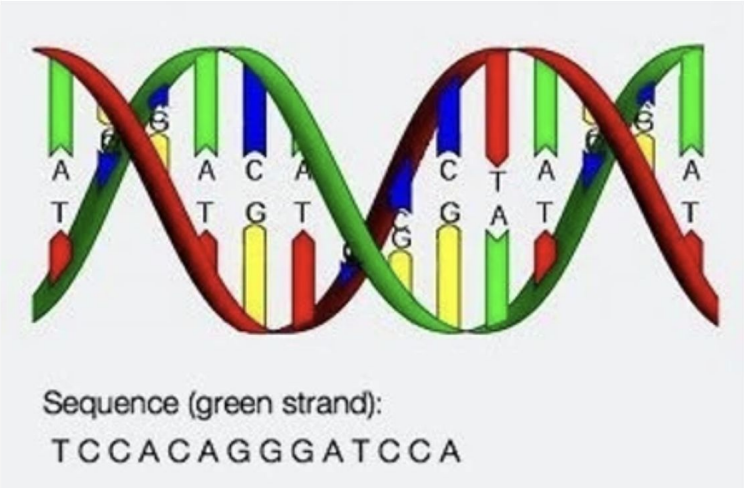
Unigram: *["a", "is", "as", "cat", "black", "white", "beautiful"]*
bigram: *[("a", "cat"), ("cat", "a"), ("black", "cat"), ("cat", "black"), ("beautiful", "cat"), ("cat", "beautiful"]*

DNA Sequence - as a string



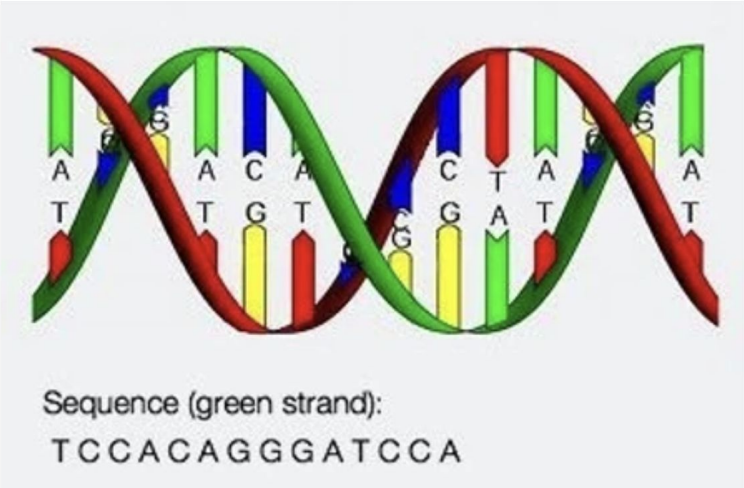


DNA Sequence - as a string




```
ENST00000435737.5
ATGTTTCGCATCACCAACATTGAGTTTCTTCCCGAATACCGACAAAAGGAGTCCAGGGAATTTCTTTCAGTGTCA
CGGACTGTGCAGCAAGTGATAAACCTGGTTTATACAACATCTGCCTTCTCCAAATTTTATGAGCAGTCTGTTGTT
GCAGATGTCAGCAACAACAAAGGCGGCCTCCTTGTCCACTTTTGGATTGTTTTTGTCATGCCACGTGCCAAAGGC
CACATCTTCTGTGAAGACTGTGTTGCCGCCATCTTGAAGGACTCCATCCAGACAAGCATCATAAACCGGACCTCT
GTGGGGAGCTTGCAGGGACTGGCTGTGGACATGGACTCTGTGGTACTAAATGAAGTCCTGGGGCTGACTCTCATT
GTCTGGATTGACTGA
390
ENST00000419127.5
ATGTTTCGCATCACCAACATTGAGTTTCTTCCCGAATACCGACAAAAGGAGTCCAGGGAATTTCTTTCAGTGTCA
CGGACTGTGCAGCAAGTGATAAACCTGGTTTATACAACATCTGCCTTCTCCAAATTTTATGAGCAGTCTGTTGTT
GCAGATGTCAGCAACAACAAAGGCGGCCTCCTTGTCCACTTTTGGATTGTTTTTGTCATGCCACGTGCCAAAGGC
CACATCTTCTGTGAAGACTGTGTTGCCGCCATCTTGAAGGACTCCATCCAGACAAGCATCATAAACCGGACCTCT
GTGGGGAGCTTGCAGGGACTGGCTGTGGACATGGACTCTGTGGTACTAAATGACAAAGGCTGCTCTCAGTACTTC
TATGCAGAGCATCTGTCTCTCCACTACCCGCTGGAGATTTCTGCAGCCTCAGGGAGGCTGATGTGTCACTTCAAG
CTGGTGGCCATAGTGGGCTACCTGATTCGTCTCTCAATCAAGTCCATCCAAATCGAAGCCGACAACTGTGTCACT
GACTCCCTGACCATTTACGACTCCCTTTTGCCCATCCGGAGCAGCATCTTGTACAGAATTTGTGAACCCACAAGA
ACATTAATGTCATTTGTTTCTACAAATAATCTCATGTTGGTGACATTTAAGTCTCCTCATATACGGAGGCTCTCA
GGAATCCGGGCATATTTTGAGGTCATTCCAGAACAAAAGTGTGAAAACACAGTGTTGGTCAAAGACATCACTGGC
TTTGAAGGGAAAATTTCAAGCCCATATTACCCGAGCTACTATCCTCCAAAATGCAAGTGTACCTGGAAATTTCAG
ACTTCTCTATCAACTCTTGGCATAGCACTGAAATTCTATAACTATTCAATAACCAAGAAGAGTATGAAAGGCTGT
GAGCATGGATGGTGGGAAATTAATGAGCACATGTACTGTGGCTCCTACATGGATCATCAGACAATTTTTCGAGTG
```

DNA Sequence - as a string




```
ENST00000435737.5
ATGTTTCGCATCACCAACATTGAGTTTCTTCCCGAATACCGACAAAAGGAGTCCAGGGAATTTCTTTCAGTGTCA
CGGACTGTGCAGCAAGTGATAAACCTGGTTTATACAACATCTGCCTTCTCCAAATTTTATGAGCAGTCTGTTGTT
GCAGATGTCAGCAACAACAAAGGCGGCCTCCTTGTCCACTTTTGGATTGTTTTTGTCATGCCACGTGCCAAAGGC
CACATCTTCTGTGAAGACTGTGTTGCCGCCATCTTGAAGGACTCCATCCAGACAAGCATCATAAACCGGACCTCT
GTGGGGAGCTTGCAGGGACTGGCTGTGGACATGGACTCTGTGGTACTAAATGACAAGGCTCCTGGGGCTGACTCTCATT
GTCTGGATTGACTGA
390
ENST00000419127.5
ATGTTTCGCATCACCAACATTGAGTTTCTTCCCGAATACCGACAAAAGGAGTCCAGGGAATTTCTTTCAGTGTCA
CGGACTGTGCAGCAAGTGATAAACCTGGTTTATACAACATCTGCCTTCTCCAAATTTTATGAGCAGTCTGTTGTT
GCAGATGTCAGCAACAACAAAGGCGGCCTCCTTGTCCACTTTTGGATTGTTTTTGTCATGCCACGTGCCAAAGGC
CACATCTTCTGTGAAGACTGTGTTGCCGCCATCTTGAAGGAGTCCATCCAGACAAGCATCATAAACCGGACCTCT
GTGGGGAGCTTGCAGGGACTGGCTGTGGACATGGACTCTGTGGTACTAAATGACAAGGCTGCTCTCAGTACTTC
TATGCAGAGCATCTGTCTCTCCACTACCGCTGGAGAATTCTGCAGCCTCAGGGAGGCTGATGTGTCACTTCAAG
CTGGTGGCCATAGTGGGCTACCTGATTCGTCTCGTGATCAAGTCCATCCAAATCGAAGCCGACAACTGTGTCACT
GACTCCCTGACCATTTACGACTTCCTTTTGCCATCCGGAGCAGCATCTTGTACAGAATTTGTGAACCCACAAGA
ACATTAATGTCATTTGTTTCTACAAATAATCTCATGTTGGTGACATTTAAGTCTCCTCATATACGGAGGCTCTCA
GGAATCCGGGCATCTTTCAGGTCATTCCAGAACAAAAGTGTGAAAACACAGTGTTGGTCAAAGACATCACTGGC
TTTGAAGGGAAAATTTCAAGCCCATATTACCCGAGCTACTATCCTCCAAAATGCAAGTGTACCTGGAAATTTCAG
ACTTCTCTATCAACTCTTGGCATAGCACTGAAATTCTATAACTATTCAATAACCAAGAAGAGTATGAAAGGCTGT
GAGCATGGATGGTGGGAAATTAATGAGCACATGTACTGTGGCTCCTACATGGATCATCAGACAATTTTTCGAGTG
```

DNA Sequence - as a string

- **Data/Feature encoding**
  - One-Hot Encoding
  - Label Encoding
- **NLP/DNA sequencing**
  - Tf-idf
  - **CountVectorizer**

Similar to what we have seen as documents!

k-mer counting!

DNA sequence as a “language”, known as k-mer counting

```
[9]  def getKmers(sequence, size=6):
         return [sequence[x:x+size].lower() for x in range(len(sequence) - size + 1)]
```

```
[62] mySeq = 'GTGCCCAGGTT'
     getKmers(mySeq, size=5)
```

```
['gtgcc', 'tgccc', 'gccca', 'cccag', 'ccagg', 'caggt', 'aggtt']
```

k-mer counting!

DNA sequence as a “language”, known as k-mer counting

```
[9]  def getKmers(sequence, size=6):
         return [sequence[x:x+size].lower() for x in range(len(sequence) - size + 1)]
```

```
[62] mySeq = 'GTGCCCAGGTT'
     getKmers(mySeq, size=5)
```

```
['gtgcc', 'tgccc', 'gccca', 'cccag', 'ccagg', 'caggt', 'aggtt']
```

Notebook presentation!