

گزارش کار تکلیف شماره ۲

در این تکلیف، با استفاده از کتابخانه hazm، سعی در انجام پردازش‌هایی ساده بر روی متون هستیم.

• گروه ۳:

- دینا امیدوار طهرانی
- پرستو باقرپور
- نسیم فانی
- محسن محمودزاده
- مریم واقعی

۱. خواندن ورودی

در ابتدای امر، با استفاده از کتابخانه docx2txt، محتویات فایل ورودی با پسوند docx را در متغیری به نام MY_TEXT ذخیره می‌کنیم.

```
MY_TEXT = docx2txt.process("paragraphs.docx")
```

۲. پیش‌پردازش ورودی

در ادامه، با فراخوانی تابع pre_process، پیش‌پردازش‌های لازم بر روی متغیر MY_TEXT را انجام می‌دهیم. انجام این کار شامل نمونه‌سازی از مازول Normalizer کتابخانه‌ی hazm و فراخوانی متد normalize این نمونه است. این پیش‌پردازش‌ها شامل پردازش‌هایی از جمله تبدیل فاصله‌ها به نیم‌فاصله‌ها، حذف کشیدگی‌های کلمات و ... است. مقدار بازگشتی متد normalize، متن نرمالیزه شده‌ی موردنظر است.

```
def pre_process(text):  
    print('normalized text is:')  
    normalizer = Normalizer()  
    text = normalizer.normalize(text)  
    print(text)  
    return text
```

```
normalized_text = pre_process(MY_TEXT)
```

normalized text is:

ابزارهایی که خیلی به گسترش این نوع از خرید کمک کرده‌اند شامل نرم افزارهای فروشگاه ساز اینترنتی، شبکه‌های اجتماعی مانند اینستاگرام و ... میشوند که باعث شدند در سال‌های آتی بیش از پیش مردم به خریدهای اینترنتی مشتاق شوند و این مدل از خرید بر دیگر روش‌های خرید ترجیح داده شود. اما سوال اصلی این است که امسال چه مقدار این رویه گسترده‌تر خواهد شد؟ طبق آمار و داده‌های به دست آمده از میزان استفاده مردم از موبایل برای خرید کالاهای مورد نیازشان این نتیجه گرفته شده است که در سال ۲۰۲۰ مادی، میزان فروش به ۲۸۴ میلیارد دلار خواهد رسید که یعنی ۴۵ درصد نسبت به سال‌های گذشته افزایش داشته است و این یعنی یک رویا برای کسب و کارهای اینترنتی. میتوان گفت که اگر این رویه در سال‌های آینده همچنان ادامه پیدا کند کسب و کارهای فیزیکی مانند فروشگاه‌ها و مغازه‌های موجود در سطح شهر ممکن است حسابی با مشکل مواجه شوند و خرید اینترنتی به اصلی‌ترین روش خرید در میان مردم تبدیل شود.

در سال‌های گذشته شبکه‌های اجتماعی به خوبی توانسته‌اند جای پای خود را در میان کاربران اینترنت سفت کنند. به عنوان مثال امروزه به سختی میتونید کسی را پیدا کنید که اینستاگرام و یا تلگرام را در موبایل خود نصب نداشته باشد و درواقع این نرم افزارها به جزی جدایی ناپذیر از دستگاه تلفن همراه مردم به ویژه مردم کشور خودمان تبدیل شده‌اند. نتایج حاصل از تحقیقات انجام شده توسط متخصصان نشان میدهد که درواقع با گذشت زمان درصد افرادی که از شبکه اجتماعی اینستاگرام استفاده میکنند به سرعت در حال افزایش است.

بنابراین این مسئله فرصت خوبی برای استفاده از این فضای پویا جهت رونق بخشیدن به کسب و کارتان است. به ویژه ازاین جهت که این شبکه امکان تبلیغات را از طریق پست‌ها و استوری‌ها برای مردم بسیار راحت‌تر نموده است. علاوه بر این افراد جوان‌تر همیشه علاقه بیشتری به شبکه‌های اجتماعی مانند اینستاگرام، یوتیوب، توئیتر، فیس بوک و این دست برنامه‌ها داشته‌اند و این گروه سنی همیشه خرید اینترنتی را به خرید حضوری ترجیح میدهند، زیرا علاوه بر راحتی این نوع از خرید، وقتی این دسته از کاربران از طریق فروشگاه‌های اینترنتی معتبر خرید خود را انجام میدهند به نوعی از اصالت کالای خود اطمینان دارند و میدانند که کالای بی که قرار است به دستشان برسد از کیفیت بالا بی برخوردار است.

۳. پردازش پاراگراف‌ها

در این گام، با استفاده از توابع `numberOfSentences` و `extractPartOfSpeech` اطلاعات مورد نیاز یعنی تعداد جملات، تعداد کلمات، تعداد فعل‌ها و تعداد اسم‌ها را برای هر پاراگراف به دست می‌آوریم. پس در ابتدا، پاراگراف‌ها را از هم جدا می‌کنیم.

در تابع `numberOfSentences`، تمامی پاراگراف‌ها را به عنوان ورودی گرفته و سپس با استفاده از ماژول `SentenceTokenizer` کتابخانه‌ی `hazm`، یک نمونه از این ماژول ساخته و با فراخوانی متد `tokenize` آن بر روی هر پاراگراف، آرایه‌ای از توکن‌ها را به دست می‌آوریم. هر یک از این توکن‌ها، یک جمله از پاراگراف هستند. برای پردازش‌های پیش رو، تعداد جملات هر پاراگراف را نیز در یک لیست به نام `numOfSentences` ذخیره می‌نماییم.

```
def numberOfSentences(paragraphs):
    ~~~~~
    numOfSentences = []
    ~~~~~
    tokenizer = SentenceTokenizer()

    i=0
    for text in paragraphs:
        sentences = tokenizer.tokenize(text)
        numOfSentences.append(len(sentences))

        i+=1
    return numOfSentences
```

در تابع `extractPartOfSpeech`، از ماژول‌های `POSTagger` و `WordTokenizer` کتابخانه `hazm` استفاده می‌کنیم. پس از نمونه‌سازی از این ماژول‌ها و با پیمایش هر پاراگراف، با استفاده از متد `tokenize` نمونه‌ی ساخته شده از `WordTokenizer`، تعداد کلمات را در آن پاراگراف به دست می‌آوریم. حال با استفاده از متد `tag` نمونه‌ی ساخته شده از `POSTagger`، توکن‌های به دست آمده از مرحله‌ی قبل را تگ می‌زنیم. حال با پیمایش بر روی این توکن‌های تگ زده شده، اگر توکن فعل بود (تگ V)، تعداد فعل‌ها را یک واحد افزایش می‌دهیم و اگر توکن اسم بود (تگ N)، به تعداد اسم‌ها یک واحد می‌افزاییم. در نهایت اطلاعات به دست آمده را در یک دیکشنری ذخیره کرده و به لیست `paragraphs_info` که دربردارنده‌ی اطلاعات مورد نظر هر پاراگراف است، اضافه می‌نماییم.

```

def extractPartOfSpeech(paragraphs, numOfSentences):
    tagger = POSTagger(model='resources/postagger.model')
    wordTokenizer = WordTokenizer()
    paragraphs_info = []
    counter = 1
    for paragraph in paragraphs:
        verbsCount = 0
        nounsCount = 0
        print('normalized paragraph'+str(counter)+':')
        print(paragraph+'\n')
        paragraphTokens = wordTokenizer.tokenize(paragraph)
        taggedTokens = tagger.tag(paragraphTokens)
        print('tagged tokens of paragraph '+str(counter)+':')
        print(str(taggedTokens)+'\n')
        wordsCount = len(taggedTokens)

```

```

        for token in taggedTokens:
            if token[1] == 'V':
                verbsCount += 1
            elif token[1] == 'N':
                nounsCount += 1
        paragraphs_info.append({'paragraphId': counter,
                               'numberOfSentences': numOfSentences[counter-1],
                               'wordsCount': wordsCount,
                               'verbsCount': verbsCount,
                               'nounsCount': nounsCount})
        counter += 1
    return paragraphs_info

```

```

paragraphs_list = normalized_text.split('\n\n') #seperate paragraphs
numOfSentences = numberOfSentences(paragraphs_list)
paragraphs_info = extractPartOfSpeech(paragraphs_list, numOfSentences)

```

normalized paragraph 1:

ایزارهایی که خیلی به گسترش این نوع از خرید کمک کرده‌اند شامل نرم افزارهای فروشگاه ساز اینترنتی، شبکه‌های اجتماعی مانند اینستاگرام و ... میشوند که باعث شدند در سال‌های آتی بیش از پیش مردم به خریدهای اینترنتی مشتاق شوند و این مدل از خرید بر دیگر روش‌های خرید ترجیح داده شود. اما سؤال اصلی این است که امسال چه مقدار این رویه گسترده‌تر خواهد شد؟ طبق آمار و داده‌های به دست‌آمده از میزان استفاده مردم از موبایل برای خرید کالاهای موردنیازشان این نتیجه گرفته شده است که در سال ۲۰۲۰ مبادی، میزان فروش به ۲۸۲ میلیارد دلار خواهد رسید که یعنی ۴۵ درصد نسبت به سال‌های گذشته افزایش داشته است و این یعنی یک رویا برای کسب وکارهای اینترنتی. میتوان گفت که اگر این رویه در سال‌های آینده همچنان ادامه پیدا کند کسبوکارهای فیزیکی مانند فروشگاه‌ها و مغازه‌های موجود در سطح شهر ممکن است حسابی با مشکل مواجه شوند و خرید اینترنتی به اصلی‌ترین روش خرید در میان مردم تبدیل شود.

tagged tokens of paragraph 1:

[['ایزارهایی', 'N', 'که', 'CONJ', 'خیلی', 'ADV', 'به', 'P', 'گسترش', 'Ne', 'این', 'DET', 'نوع', 'N', 'از', 'P', 'خرید', 'N', 'کمک', 'N', 'کرده', 'u200c', 'اند', 'V', 'شامل', 'AJe', 'نرم', 'AJe', 'افزار', 'u200c', 'های', 'Ne', 'فروشگاه', 'Ne', 'ساز', 'Ne', 'اینترنتی', 'AJ', 'و', 'PUNC', 'شبکه', 'u200c', 'های', 'Ne', 'اجتماعی', 'AJ', 'مانند', 'ADVe', 'اینستاگرام', 'N', 'و', 'CONJ', 'و', 'PUNC', 'میشوند', 'V', 'که', 'CONJ', 'باعث', 'AJ', 'شدند', 'V', 'در', 'P', 'سال', 'u200c', 'های', 'Ne', 'آتی', 'AJ', 'بیش', 'AJ', 'از', 'P', 'پیش', 'ADVe', 'مردم', 'N', 'به', 'P', 'خریده‌ای', 'Ne', 'اینترنتی', 'AJ', 'مشتاق', 'AJ', 'شوند', 'V', 'و', 'CONJ', 'این', 'DET', 'مدل', 'N', 'از', 'P', 'خرید', 'N', 'بر', 'P', 'دیگر', 'PRO', 'روش', 'u200c', 'های', 'Ne', 'خرید', 'N', 'ترجیح', 'N', 'داده‌شود', 'V', 'و', 'PUNC', 'اما', 'CONJ', 'سؤال', 'Ne', 'اصلی', 'AJ', 'این', 'PRO', 'است', 'V', 'که', 'CONJ', 'امسال', 'ADV', 'چه', 'DET', 'مقدار', 'N', 'این', 'DET', 'رویه', 'N', 'گسترده‌تر', 'AJ', 'خواهد‌شد', 'V', '؟', 'PUNC', 'خلق', 'Ne', 'آمار', 'N', 'و', 'CONJ', 'داده‌های', 'N', 'به', 'P', 'دست‌آمده', 'N', 'از', 'P', 'میزان', 'Ne', 'استفاده', 'Ne', 'مردم', 'N', 'از', 'P', 'موبایل', 'N', 'برای', 'Pe', 'خرید', 'Ne', 'کالاهای', 'Ne', 'موردنیازشان', 'Ne', 'این', 'DET', 'نتیجه', 'N', 'گرفته‌شده‌است', 'V', 'که', 'CONJ', 'در', 'P', 'سال', 'Ne', 'NUM', '۲۰۲۰', 'مبادی', 'N', 'و', 'PUNC', 'میزان', 'Ne', 'فروش', 'N', 'به', 'P', 'NUM', '۲۸۲', 'میلیارد', 'NUM', 'دلار', 'N', 'خواهد‌رسید', 'V', 'که', 'CONJ', 'یعنی', 'CONJ', 'NUM', '۴۵', 'درصد', 'N', 'نسبت', 'N', 'به', 'P', 'سال', 'u200c', 'های', 'Ne', 'گذشته', 'ADVe', 'افزایش', 'N', 'و', 'CONJ', 'این', 'PRO', 'یعنی', 'CONJ', 'یک', 'NUM', 'رویا', 'N', 'برای', 'Pe', 'کسب', 'Ne', 'وکارهای', 'Ne', 'اینترنتی', 'AJ', 'و', 'PUNC', 'میتوان', 'V', 'گفت', 'V', 'که', 'CONJ', 'اگر', 'CONJ', 'این', 'DET', 'رویه', 'N', 'در', 'P', 'سال', 'u200c', 'های', 'Ne', 'آینده', 'N', 'همچنان', 'ADVe', 'ادامه', 'N', 'پیدا', 'AJ', 'کند', 'V', 'کسبوکارهای', 'Ne', 'فیزیکی', 'AJ', 'مانند', 'ADVe', 'فروشگاه', 'u200c', 'ها', 'N', 'و', 'CONJ', 'مغازه', 'u200c', 'های', 'Ne', 'موجود', 'AJ', 'در', 'P', 'شهر', 'N', 'ممکن', 'AJ', 'است', 'V', 'حسابی', 'N', 'با', 'P', 'مشکل', 'N', 'مواجه', 'AJ', 'شوند', 'V', 'و', 'CONJ', 'خرید', 'Ne', 'اینترنتی', 'AJ', 'به', 'P', 'اصلی', 'u200c', 'ترین', 'AJ', 'روش', 'Ne', 'خرید', 'N', 'در', 'P', 'میان', 'Ne', 'مردم', 'N', 'تبدیل', 'N', 'شود', 'V', 'و', 'PUNC', 'و']]

normalized paragraph 2:

در سال‌های گذشته شبکه‌های اجتماعی به خوبی توانسته‌اند جای پای خود را در میان کاربران اینترنت سفت کنند. به عنوان مثال امروزه به سختی میتوانید کس را پیدا کنید که اینستاگرام و یا تلگرام را در موبایل خود نصب نداشته باشد و درواقع این نرم افزارها به جزئی جدایی ناپذیر از دستگاه تلفن همراه مردم به ویژه مردم کشور خودمان تبدیل شده‌اند. نتایج حاصل از تحقیقات انجام شده توسط متخصصان نشان میدهد که درواقع با گذشت زمان درصد افرادی که از شبکه‌ی اجتماعی اینستاگرام استفاده میکنند به سرعت در حال افزایش است.

tagged tokens of paragraph 2:

[['در', 'P', 'سال‌های', 'Ne', 'گذشته', 'ADVe', 'شبکه‌های', 'Ne', 'اجتماعی', 'AJ', 'به', 'P', 'خوبی', 'N', 'توانسته', 'u200c', 'اند', 'V', 'جای', 'Ne', 'پای', 'Ne', 'خود', 'PRO', 'را', 'POSTP', 'در', 'P', 'میان', 'Ne', 'کاربران', 'Ne', 'اینترنت', 'N', 'سفت', 'AJ', 'کنند', 'V', 'و', 'PUNC', 'به', 'P', 'عنوان', 'Ne', 'مثال', 'N', 'امروزه', 'ADVe', 'به', 'P', 'سختی', 'N', 'میتوانید', 'V', 'کسی', 'N', 'را', 'POSTP', 'پیدا', 'AJ', 'کنید', 'V', 'که', 'CONJ', 'اینستاگرام', 'N', 'و', 'CONJ', 'با', 'CONJ', 'تلگرام', 'N', 'را', 'POSTP', 'در', 'P', 'موبایل', 'Ne', 'خود', 'PRO', 'نصب', 'N', 'نداشته‌باشد', 'V', 'و', 'CONJ', 'درواقع', 'CONJ', 'این', 'DET', 'نرم', 'AJ', 'افزارها', 'N', 'به', 'P', 'جزئی', 'N', 'جدایی', 'Ne', 'ناپذیر', 'AJ', 'از', 'P', 'دستگاه', 'Ne', 'تلفن', 'Ne', 'همراه', 'AJe', 'مردم', 'N', 'به', 'P', 'ویژه', 'AJ', 'مردم', 'Ne', 'کشور', 'Ne', 'خودمان', 'PRO', 'تبدیل', 'N', 'شده', 'u200c', 'اند', 'V', 'و', 'PUNC', 'نتایج', 'Ne', 'حاصل', 'N', 'از', 'P', 'تحقیقات', 'N', 'انجام', 'N', 'شده', 'V', 'توسط', 'Pe', 'متخصصان', 'N', 'نشان', 'N', 'میبدهد', 'V', 'که', 'CONJ', 'درواقع', 'CONJ', 'با', 'P', 'گذشت', 'Ne', 'زمان', 'N', 'درصد', 'Ne', 'افراد', 'N', 'که', 'CONJ', 'از', 'P', 'شبکه', 'u200c', 'ای', 'Ne', 'اجتماعی', 'AJe', 'اینستاگرام', 'N', 'استفاده', 'N', 'میکنند', 'V', 'به', 'P', 'سرعت', 'N', 'در', 'P', 'حال', 'Ne', 'افزایش', 'N', 'است', 'V', 'و', 'PUNC', 'و']]

بنابراین این مسئله فرصت خوبی برای استفاده از این فضای پویا جهت رونق بخشیدن به کسب و کارتان است، به ویژه از این جهت که این شبکه امکان تبلیغات را از طریق پستها و استوریها برای مردم بسیار راحتتر نموده است. علاوه بر این افراد جوانتر همیشه علاقه بیشتری به شبکه‌های اجتماعی مانند اینستاگرام، یوتیوب، توئیتر، فیس بوک و این دست برنامه‌ها داشته‌اند و این گروه سنی همیشه خرید اینترنتی را به خرید حضوری ترجیح می‌دهند، زیرا علاوه بر راحتی این نوع از خرید، وقتی این دست کاربران از طریق فروشگاه‌های اینترنتی معتبر خرید خود را انجام می‌دهند به نوعی از اسالت کالاوی خود اطمینان دارند و میدانند که کالا بهی که قرار است به دستشان برسد از کیفیت بالا بی‌بخوردار است.

[¹ «بنا براین» (CONJ)، «این» (DET)، «مسئله» (N)، «فرست» (Ne)، «خوبی» (AJ)، «برای» (Pe)، «تفاده» (N)، «آزار» (P)، «این» (N)، «فصلی» (Ne)، «بویا» (AJ)، «جهت» (Pe)، «رونی» (N)، «بخشیدن» (N)، «به» (N)، «کس» (Ne)، «و کارتان» (N)، «است» (V)، «این» (N)، «پنج» (P)، «روز» (AJ)، «آزار» (DET)، «جهت» (N)، «که» (CONJ)، «این» (DET)، «شک» (N)، «امکان» (N)، «Ne» (N)، «تعلیقات» (N)، «آزار» (POSTP)، «آزار» (P)، «طریق» (Ne)، «یستفا» (N)، «و» (CONJ)، «استوریا» (N)، «برای» (Pe)، «مردم» (N)، «بسیار» (AJ)، «راحتی» (ADJ)، «نموده» (N)، «این» (DET)، «علاوه» (N)، «بر» (P)، «این» (DET)، «افراد» (N)، «جوانتر» (AJ)، «همیشه» (ADV)، «علاقه» (200c)، «Ne» (N)، «یستقر» (AJ)، «به» (P)، «شک» (200c)، «Ne» (N)، «اجتماعی» (AJ)، «مانند» (ADVe)، «اینستاگرام» (N)، «این» (PUNC)، «یوتیوب» (N)، «این» (PUNC)، «تویتر» (N)، «این» (PUNC)، «فیس» (Ne)، «یوت» (N)، «و» (CONJ)، «این» (DET)، «دست» (CL)، «برنامه» (200c)، «این» (N)، «داشته» (200c)، «این» (V)، «و» (CONJ)، «این» (DET)، «گروه» (Ne)، «سعی» (AJ)، «همیشه» (ADV)، «خرید» (Ne)، «این» (AJ)، «اینترنتی» (AJ)، «خرید» (P)، «به» (P)، «خرید» (Ne)، «خوبی» (AJ)، «ترجیح» (N)، «میدهند» (V)، «این» (PUNC)، «آزیرا» (CONJ)، «عاه» (N)، «بر» (P)، «راحتی» (Ne)، «این» (DET)، «و» (N)، «آزار» (P)، «خرید» (N)، «این» (PUNC)، «وقتی» (CONJ)، «این» (DET)، «دسته» (N)، «کاربران» (N)، «آزار» (P)، «طریق» (Ne)، «روشگاه» (200c)، «Ne» (N)، «اینترنتی» (AJ)، «اعتبار» (AJ)، «خرید» (Ne)، «خود» (PRD)، «آزار» (POSTP)، «انجام» (N)، «میدهند» (V)، «رونی» (N)، «آزار» (P)، «اصالت» (Ne)، «کالا» (N)، «قی» (Ne)، «خود» (PRD)، «امضای» (N)، «دارند» (V)، «و» (CONJ)، «میدانند» (V)، «که» (CONJ)، «کالا» (N)، «بی» (N)، «که» (CONJ)، «قران» (N)، «است» (V)، «به» (P)، «دستشان» (N)، «برسد» (V)، «آزار» (P)، «کیفیت» (Ne)، «کالا» (AJ)، «بی» (N)، «برخورداری» (AJ)، «است» (V)، «این» (PUNC)]

```
{['paragraphId': '1', 'numberOfSentences': 4, 'wordsCount': 161, 'verbsCount': 16, 'nounsCount': 35], ['paragraphId': '2', 'numberOfSentences': 3, 'wordsCount': 94, 'verbsCount': 10, 'nounsCount': 23], ['paragraphId': '3', 'numberOfSentences': 2, 'wordsCount': 130, 'verbsCount': 10, 'nounsCount': 33]}
```

تعداد اسم‌ها	تعداد فعل‌ها	تعداد کل کلمات	تعداد جملات	
۳۵	۱۶	۱۶۱	۴	پاراگراف ۱
۲۳	۱۰	۹۴	۳	پاراگراف ۲
۳۳	۱۰	۱۳۰	۲	پاراگراف ۳

در این گام، به کمک به متغیری به نام counter، ۵ جمله اول متن را انتخاب می‌کنیم. در لیست allSentences نیز تمامی جملات متن را

WordTokenizer، توکن‌های هر جمله را به دست آورده و برای هر توکن، در صورتی که کلمه یا فعل باشد، به کمک نمونه ساخته شده از

ماژول Stemmer کتابخانه‌ی hazm، ریشه‌ی آن توکن را به دست می‌آوردیم و حاصل را در `stem_list` ذخیره می‌کنیم.


```
def fivesentences(paragraphs):
    # print('number of sentences for each paragraph:')
    global sen
    tokenizer = SentenceTokenizer()
    tokenizer1= WordTokenizer()
    stemmer = Stemmer()
    allSentences = []
    for text in paragraphs:
        sentences = tokenizer.tokenize(text) #find sentences in paragraph
        for sen in sentences:
            allSentences.append(sen)

    counter = 1
    for sentence in allSentences:
        if(counter != 6):
            print('sentence '+str(counter)+':')
            print(sentence)
            print('stems of sentence '+str(counter)+':')
            stem_list = []
            x = tokenizer1.tokenize(sentence) #find tokens of one sentence
            for element in x: #for each token of sentence wana find stem
                if "." not in element and "," not in element and "..." not in element:

                    if not element.isdigit():
                        el = stemmer.stem(element)
                        stem_list.append(el)
            print(str(stem_list)+'\n')
            counter += 1
```

sentence 1:

ابزارهایی که خیلی به گسترش این نوع از خرید کمک کرده‌اند شامل نرم افزارهای فروشگاه ساز اینترنتی، شبکه‌های اجتماعی مانند اینستاگرام و ... میشوند که باعث شدند در سال‌های آتی بیش از پیش مردم به خریدهای اینترنتی مشتاق شوند و این مدل از خرید بر دیگر روش‌های خرید ترجیح داده شود.

stems of sentence 1:

['ابزار', 'که', 'خیلی', 'به', 'گسترش', 'این', 'نوع', 'از', 'خرید', 'کمک', 'کرده', 'شامل', 'نرم', 'افزار', 'فروشگاه', 'ساز', 'اینترنت', 'شبکه', 'اجتماعی', 'مانند', 'اینستاگرام', 'و', '...', 'میشوند', 'که', 'باعث', 'شدند', 'در', 'سال', 'آتی', 'بیش', 'از', 'پیش', 'مردم', 'به', 'خرید', 'اینترنت', 'مشتاق', 'شدند', 'و', 'این', 'مدل', 'از', 'خرید', 'بر', 'دیگر', 'روش', 'خرید', 'ترجیح', 'داده‌شود']

sentence 2:

اما سؤال اصلی این است که امسال چه مقدار این رویه گسترده‌تر خواهد شد؟

stems of sentence 2:

['اما', 'سؤال', 'اصلی', 'این', 'است', 'که', 'امسال', 'چه', 'مقدار', 'این', 'رویه', 'گسترده', 'خواهد', 'شد', '؟']

sentence 3:

طبق آمار و داده‌های به دست‌آمده از میزان استفاده مردم از موبایل برای خرید کالاهای موردنیازشان این نتیجه گرفته شده است که در سال ۲۰۲۰ مبادی، میزان فروش به ۲۸۴ میلیارد دلار خواهد رسید که یعنی ۳۵ درصد نسبت به سال‌های گذشته افزایش داشته است و این یعنی یک رویا برای کسب وکارهای اینترنتی.

stems of sentence 3:

['طبق', 'آمار', 'و', 'داده', 'به', 'دست‌آمده', 'از', 'میزان', 'استفاده', 'مردم', 'از', 'موبایل', 'برای', 'خرید', 'کالا', 'موردنیاز', 'این', 'نتیجه', 'گرفته‌شده', 'است', 'که', 'در', 'سال', 'سال', 'مبادی', 'میزان', 'فروش', 'به', 'به', 'میلیارد', 'دلار', 'خواهد', 'رسید', 'که', 'یعنی', 'این', 'درصد', 'نسب', 'به', 'سال', 'گذشته', 'افزایش', 'داشته', 'است', 'و', 'این', 'یعنی', 'یک', 'رویا', 'برای', 'کسب', 'وکار', 'اینترنت']

sentence 4:

میتوان گفت که اگر این رویه در سال‌های آینده همچنان ادامه پیدا کند کسب‌وکارهای فیزیکی مانند فروشگاه‌ها و مغازه‌های موجود در سطح شهر ممکن است حسابی با مشکل مواجه شوند و خرید اینترنتی به اصلی‌ترین روش خرید در میان مردم تبدیل شود.

stems of sentence 4:

['میتوان', 'گفت', 'که', 'اگر', 'این', 'رویه', 'در', 'سال', 'آینده', 'همچنان', 'ادامه', 'پیدا', 'کند', 'کسب‌وکار', 'فیزیکی', 'مانند', 'فروشگاه', 'و', 'مغازه', 'موجود', 'در', 'سطح', 'شهر', 'ممکن', 'است', 'حسابی', 'با', 'مشکل', 'مواجه', 'شوند', 'و', 'خرید', 'اینترنت', 'به', 'اصلی', 'روش', 'خرید', 'در', 'میان', 'مردم', 'تبدیل', 'شود']

sentence 5:

در سالیان گذشته شبکه‌های اجتماعی به خوبی توانسته‌اند جای پای خود را در میان کاربران اینترنت سفت کنند.

stems of sentence 5:

['در', 'سال', 'گذشته', 'شبکه', 'اجتماعی', 'به', 'خوب', 'توانسته', 'جای', 'پای', 'خود', 'را', 'در', 'میان', 'کاربران', 'اینترنت', 'سفت', 'کنند']

از مواردی که ریشه‌یابی آنها به درستی در مرحله قبل ظاهر شده‌اند، می‌توان به موارد زیر اشاره کرد:

۱. توانسته‌اند <--- توانست#توان

۲. نداشته‌باشند <--- داشت#دار

۳. برنامه‌ها <--- برنامه

۴. متخصصان <--- متخصص