



دانشکده مهندسی کامپیوتر

دانشگاه صنعتی شریف

پروژه درس سیستم‌های پشتیبان تصمیم‌گیری

آشنایی با تحلیل نظرات کاربران شبکه‌های اجتماعی
برای پشتیبانی از تصمیم‌گیری

استاد درس: دکتر جعفر حبیبی

شماره گروه: ۳

نام و نام خانوادگی: امیررضا شفاعت

شماره دانشجویی: ۹۶۲۰۹۵۶۹

نیمسال دوم ۱۳۹۶-۱۳۹۷

فهرست مطالب

۱ گام اول..... ۱

- ۱/۱ مقاله اول: طراحی یک سیستم پشتیبان تصمیم‌گیری مبتنی بر AHP، به منظور انتخاب
تأمین‌کننده در صنعت حمل و نقل ۱
- ۱/۱/۱ مقدمه ۱
- ۱/۱/۲ اهمیت مسئله ۱
- ۱/۱/۳ روش AHP ۲
- ۱/۱/۴ تعیین شاخص‌ها و انتخاب تأمین‌کننده به روش AHP ۳
- ۱/۱/۵ تحلیل حساسیت ۵

۱/۲ مقاله دوم: یک رویکرد جدید برای طراحی یک سیستم پشتیبان تصمیم‌گیری بالینی

- مبتنی بر هستی‌شناسی ۶
- ۱/۲/۱ مقدمه ۶
- ۱/۲/۲ دیابت و مالاریا ۷
- ۱/۲/۳ مدیریت دانش ۷
- ۱/۲/۴ مهندسی دانش ۸
- ۱/۲/۵ جمع‌بندی ۸

۱/۳ روش‌ها و ابزارهای تحلیل داده‌های حجیم..... ۸

- ۱/۳/۱ روش شماره ۱: انبارداده و مفاهیم مرتبط ۹
- ۱/۳/۲ ابزار شماره ۲: Apache Cassandra ۱۱
- ۱/۳/۳ ابزار شماره ۳: کتابخانه fastText ۱۳
- ۱/۳/۴ ابزار شماره ۴: NLTK ۱۴
- ۱/۳/۵ روش شماره ۵: رگرسیون خطی ۱۶
- ۱/۳/۶ روش شماره ۶: الگوریتم C4.5 ۱۶

۲ گام دوم..... ۱۹

۳ گام سوم..... ۳۱

۳۱	گزارش شماره ۱
۳۲	گزارش شماره ۲
۳۳	گزارش شماره ۳
۳۴	گزارش شماره ۴
۳۶	گزارش شماره ۵
۳۷	گزارش شماره ۶
۳۸	گزارش شماره ۷
۳۹	گزارش شماره ۸
۴۱	گزارش شماره ۹
۴۲	گزارش شماره ۱۰
۴۴	گزارش شماره ۱۱
۴۴	گزارش شماره ۱۲
۴۵	ضمایم
۴۵	ضمیمه شماره ۱: هرم DIKW و تفاوت سطوح مختلف
۴۷	مراجع

فهرست تصاویر

شکل ۱-۱	سلسله مراتب روش AHP	۴
شکل ۲-۱	مدیریت دانش در سیستم حفاظت از سلامت	۸
شکل ۳-۱	مدل سنتی انبارداده	۱۱
شکل ۴-۱	شمای داده‌ها در پایگاه‌داده‌ی آپاچی کاساندرا	۱۲
شکل ۵-۱	معماری کلان پایگاه‌داده‌ی آپاچی کاساندرا	۱۳
شکل ۶-۱	نمونه‌ای فرضی از نمایش لغت با بردار	۱۴
شکل ۷-۱	تحلیل احساس متن به کمک NLTP	۱۵
شکل ۸-۱	داده‌های ورودی برای الگوریتم C4.5	۱۸
شکل ۹-۱	درخت تصمیم خروجی از الگوریتم C4.5	۱۸
شکل ۱-۴	هرم DIKW	۴۶
شکل ۲-۴	جایگاه سطوح مختلف DIKW در محور زمان	۴۶

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

گام اول

در بخش اول این گام به بررسی دو مقاله در زمینه‌ی سیستم‌های پشتیبان تصمیم‌گیری می‌پردازیم:

۱/۱ مقاله اول: طراحی یک سیستم پشتیبان تصمیم‌گیری مبتنی بر AHP، به منظور انتخاب تأمین‌کننده در صنعت حمل و نقل [1]

۱/۱/۱ مقدمه

هدف این مقاله عرضه‌ی یک مدل پشتیبان تصمیم‌گیری برای انتخاب تأمین‌کنندگان صنعت حمل‌ونقل در کشور پاکستان و به کمک روش AHP است. به علاوه پس از اعمال روش AHP و تعیین تأمین‌کننده مناسب، بر روی نتیجه آنالیز حساسیت صورت گرفته تا استواری پاسخ مشخص گردد. باید توجه داشت که این مسئله یک مسئله‌ی تصمیم‌گیری چند شاخصه^۱ می‌باشد.

در مرحله اول این مدل، شاخص‌های اصلی تأثیرگذار بر انتخاب تأمین‌کننده تعیین می‌گردد. این عمل بر مبنای نظرات متخصصان حوزه حمل‌ونقل و همچنین مرور پژوهش‌های قبلی در این زمینه صورت می‌گیرد. در مرحله دوم، نوبت به تعیین زیرشاخص‌های تأثیرگذار بر مسئله با توجه شاخص‌های اصلی تعیین شده در مرحله اول می‌رسد. در نهایت به کمک نرم افزار Expert Choice بر روی نتیجه تحلیل حساسیت صورت می‌گیرد.

این تحقیق به فعالان حوزه‌ی حمل و نقل کمک نموده تا بتوانند امر تصمیم‌گیری و انتخاب تأمین‌کننده را به صورت علمی و مبتنی بر یک روش استوار انجام دهند. همچنین با توجه به نتایج تحلیل حساسیت، می‌توان دریافت که با چه تغییراتی در شاخص‌های اصلی، تأمین‌کننده می‌بایست تغییر نماید. به این ترتیب مسئله‌ی مهم و پیچیده‌ی انتخاب تأمین‌کننده، به کمک روش AHP به یک سلسله‌مراتب ساده‌تر تجزیه شده و مسئله‌ی انتخاب تأمین‌کننده برای مدیران تسهیل می‌گردد.

۱/۱/۲ اهمیت مسئله

با توجه به نقش مهم تأمین‌کنندگان در زنجیره تأمین کلی سازمان‌ها، انتخاب تأمین‌کننده مناسب بسیار حیاتی می‌باشد. تولید وسایل نقلیه در کشور پاکستان یک بخش رو به رشد و فعال در سال‌های متمادی بوده است. هر چند تولید ۱۰۰ تا ۱۷۰ هزار اتومبیل به صورت سالیانه در مقایسه با صنایع برجسته و پیش‌تاز در جهان کوچک محسوب می‌گردد. البته در سال‌های اخیر برخی از شرکت‌های بزرگ

¹ Multi-Criteria Decision Making (MCDM) Problem

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

تولید اتومبیل اقدام ایجاد شرکت‌های محلی در پاکستان با سرمایه‌گذاری مشترک نموده‌اند. شایان ذکر است که حدود ۵۵۰۰ تا ۶۰۰۰ شغل مستقیم توسط این صنعت در پاکستان ایجاد گردیده است.

صنعت تولید وسایل نقلیه در پاکستان دومین صنعت بزرگ محسوب شده و تعداد زیادی کارخانه در پاکستان به تولید قطعات مورد نیاز تولید وسایل نقلیه می‌پردازند. این صنعت تأثیر زیادی در رشد اقتصادی کشور پاکستان ایفا می‌کند.

روش AHP ۱/۱/۳

در روش AHP در مرحله‌ی اول لازم است تمام شاخص‌های تصمیم‌گیری معین گردد. پس از تعیین شاخص‌های تصمیم‌گیری، می‌بایست ماتریس مقایسه‌ی زوجی شاخص‌ها ایجاد گردد. این ماتریس یک ماتریس مربعی متقارن با قطر اصلی یک می‌باشد که در آن بین هر دو شاخص مسئله یک مقایسه انجام شده و ارجحیت آن با توجه به جدول زیر مشخص گردیده است:

امتیاز	میزان اهمیت
۱	اهمیت شاخص i با j یکسان است.
۳	اهمیت شاخص i به میزان کمی از j بیشتر است.
۵	اهمیت شاخص i به میزان زیادی از j بیشتر است.
۷	اهمیت شاخص i به میزان خیلی زیادی از j بیشتر است.
۹	شاخص i به طور کامل بر شاخص j ارجحیت دارد
❖ از اعداد زوج برای شرایط میانی استفاده می‌گردد.	

به طور مثال یک نمونه از ماتریس مقایسه زوجی برای یک مسئله تصمیم‌دارای ۳ شاخص، می‌تواند به صورت زیر باشد:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

در مرحله بعد می‌بایست ماتریس نرمال گردد. برای نرمال شدن ماتریس A، لازم است که تمام مقادیر به مجموع ستون مربوطه تقسیم گردند:

$$N = \begin{bmatrix} \frac{a_{11}}{a_{11} + a_{21} + a_{31}} & \frac{a_{12}}{a_{12} + a_{22} + a_{32}} & \frac{a_{13}}{a_{13} + a_{23} + a_{33}} \\ \frac{a_{21}}{a_{11} + a_{21} + a_{31}} & \frac{a_{22}}{a_{12} + a_{22} + a_{32}} & \frac{a_{23}}{a_{13} + a_{23} + a_{33}} \\ \frac{a_{31}}{a_{11} + a_{21} + a_{31}} & \frac{a_{32}}{a_{12} + a_{22} + a_{32}} & \frac{a_{33}}{a_{13} + a_{23} + a_{33}} \end{bmatrix}$$

در ادامه می‌بایست بردار وزن شاخص‌ها (W) تشکیل گردد. این بردار از میانگین سطری ماتریس نرمال‌شده‌ی ماتریس مقایسه زوجی (N) حاصل می‌شود:

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

$$W = \begin{bmatrix} \frac{n_{11} + n_{12} + n_{13}}{3} \\ \frac{n_{21} + n_{22} + n_{23}}{3} \\ \frac{n_{31} + n_{32} + n_{33}}{3} \end{bmatrix}$$

در نهایت می‌بایست بردار $A \times W$ محاسبه گردد. این بردار شاخص نهایی ما برای ارزیابی تصمیم‌های موجود می‌باشد.

به منظور بررسی میزان استواری نتایج، لازم است مقدار ویژه بردار $A \times W$ محاسبه گردد. این عدد برابر با جمع مقادیر درایه‌های بردار $A \times W$ می‌باشد:

$$\lambda_{max} = \sum A \times W$$

در صورتی که میزان λ_{max} به تعداد شاخص‌های مسئله (در این مثال عدد ۳) نزدیک باشد، جواب استوار محسوب می‌گردد. به این منظور سنج‌های مربوط به استواری به صورت زیر تعریف می‌گردد:

$$Consistency Index (CI) = \frac{\lambda_{max} - n}{n - 1}$$

$$Random Inconsistency (RI) = \frac{1.98 \times (n - 2)}{n}$$

$$Consistency Ratio (CR) = \frac{CI}{RI}$$

* n تعداد شاخص‌های مسئله تصمیم‌گیری است.

** در صورتی که مقدار CR ، کمتر یا مساوی ۰/۱ باشد، میزان استواری قابل قبول است.

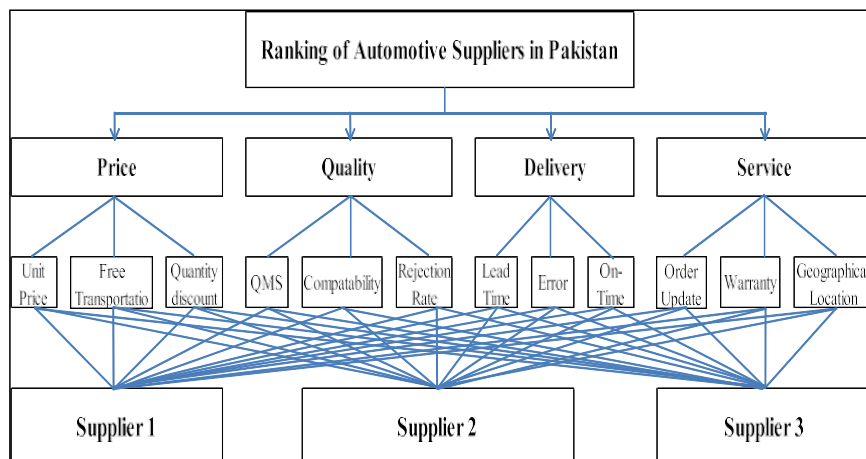
۱/۱/۴ تعیین شاخص‌ها و انتخاب تأمین‌کننده به روش AHP

با توجه به بررسی پژوهش‌های پیشین انجام شده و همچنین نظر متخصصان حوزه تولید و سایل نقلیه در کشور پاکستان، شاخص‌های اصلی انتخاب یک تأمین‌کننده عبارتند از:

۱. قیمت
۲. کیفیت
۳. سطح خدمت‌رسانی
۴. تحویل

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

همچنین برای هر کدام از شاخص‌های اصلی، تعدادی زیرشاخص نیز در نظر گرفته شده است. این موارد در نمودار زیر نمایش داده شده است:



شکل ۱-۱ سلسله مراتب روش AHP

محاسبات زیادی برای حصول نتیجه نهایی در نرم افزار Expert Choice انجام گردیده است که فقط موارد مربوط به شاخص‌های اصلی در ادامه نمایش داده شده است:

جدول ۱-۱ مقایسه زوجی شاخص‌های اصلی

	Delivery	Price	Quality	Service
Delivery	1	$\frac{1}{3}$	$\frac{1}{2}$	2
Price	3	1	2	4
Quality	2	$\frac{1}{2}$	1	3
Service	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{3}$	1

بنابر این ماتریس نرمال شده و ماتریس وزن شاخص‌ها به صورت زیر خواهد بود:

$$N = \begin{bmatrix} 0.15 & 0.16 & 0.13 & 0.20 \\ 0.46 & 0.48 & 0.52 & 0.40 \\ 0.31 & 0.24 & 0.26 & 0.30 \\ 0.08 & 0.12 & 0.01 & 0.10 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.16 \\ 0.47 \\ 0.28 \\ 0.10 \end{bmatrix}$$

در ادامه می‌بایست بردار $A \times W$ را محاسبه نمود:

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

$$AW = \begin{bmatrix} 0.65 \\ 1.89 \\ 1.12 \\ 0.38 \end{bmatrix}$$

همچنین سنجه‌های استواری به صورت زیر محاسبه می‌گردد:

$$\lambda_{max} = 4.04$$

$$CI = 0.0133$$

$$RI = 0.99$$

$$CR = 0.01$$

با توجه به اینکه مقدار CR از ۰/۱ کمتر می‌باشد، نتایج استوار محسوب می‌گردد.

به طریق مشابه بردار وزن زیرشاخص‌ها نیز معین گردیده و پس از امتیازدهی به هر یک از تأمین‌کنندگان بر اساس شاخص‌های اصلی و زیرمجموعه، با ضرب بردار امتیازات در بردار وزن، امتیاز نهایی هر تأمین‌کننده معین می‌گردد. تمام محاسبات مربوطه توسط نرم افزار Expert Choice انجام گرفته و اولویت نهایی انتخاب تأمین‌کنندگان به صورت زیر تعیین گردیده است:

(۱) تأمین‌کننده شماره ۲

(۲) تأمین‌کننده شماره ۱

(۳) تأمین‌کننده شماره ۳

۱/۱/۵ تحلیل حساسیت

تحلیل حساسیت مسئله با توجه به هر کدام از شاخص‌های اصلی به صورت زیر خواهد بود:

۱/۱/۵/۱ تحلیل حساسیت با توجه به قیمت:

❖ افزایش تأثیر قیمت:

رتبه‌بندی تأمین‌کنندگان از (۳-۱-۲) به (۱-۳-۲) تغییر خواهد کرد در صورتی که تأثیر قیمت ۶۸٪، کیفیت ۱۷٪، تحویل ۹٪ و خدمت‌رسانی ۶٪ تغییر پیدا کند.

❖ کاهش تأثیر قیمت:

همچنین رتبه‌بندی از (۳-۱-۲) به (۳-۲-۱) تغییر خواهد کرد در صورتی که تأثیر قیمت ۱۳٪، کیفیت ۳۷٪، تحویل ۲۱٪ و خدمت‌رسانی ۱۱٪ تغییر پیدا کند.

۱/۱/۵/۲ تحلیل حساسیت با توجه به کیفیت:

❖ افزایش تأثیر کیفیت:

رتبه‌بندی تأمین‌کنندگان از (۳-۱-۲) به (۳-۲-۱) تغییر خواهد کرد در صورتی که تأثیر قیمت ۲۳٪، کیفیت ۶۶٪، تحویل ۶٪ و خدمت‌رسانی ۵٪ تغییر پیدا کند.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

❖ کاهش تأثیر کیفیت:

کاهش تأثیر کیفیت، رتبه‌بندی تأمین‌کنندگان را تغییر نخواهد داد.

۱/۱/۵/۳ تحلیل حساسیت با توجه به تحویل:

❖ افزایش تأثیر تحویل:

رتبه‌بندی تأمین‌کنندگان از (۳-۱-۲) به (۳-۲-۱) تغییر خواهد کرد در صورتی که تأثیر قیمت ۴۰٪، کیفیت ۲۴٪، تحویل ۳۰٪ و خدمت‌رسانی ۸٪ تغییر پیدا کند.

❖ کاهش تأثیر تحویل:

کاهش تأثیر تحویل، رتبه‌بندی تأمین‌کنندگان را تغییر نخواهد داد.

۱/۱/۵/۴ تحلیل حساسیت با توجه به خدمت‌رسانی:

❖ افزایش تأثیر خدمت‌رسانی:

افزایش تأثیر خدمت‌رسانی، رتبه‌بندی تأمین‌کنندگان را تغییر نخواهد داد.

❖ کاهش تأثیر خدمت‌رسانی:

کاهش تأثیر خدمت‌رسانی، رتبه‌بندی تأمین‌کنندگان را تغییر نخواهد داد.

۱/۲ مقاله دوم: یک رویکرد جدید برای طراحی یک سیستم پشتیبان تصمیم‌گیری

بالینی مبتنی بر هستی‌شناسی [2]

۱/۲/۱ مقدمه

برنامه‌های دارویی یکی از حیاتی‌ترین سامانه‌های موجود می‌باشند. هر گونه خطای تصمیم‌گیری توسط پزشک و یا سیستم پشتیبان تصمیم‌گیری می‌تواند منجر به مرگ بیمار شود. باید توجه داشت که اطلاعات دارویی یکی از مهم‌ترین داده‌هایی هستند که توسط سیستم‌های پشتیبان تصمیم‌گیری پردازش می‌شوند. امروزه به کمک الگوریتم‌های هوش مصنوعی سامانه‌های مکملی برای کمک به تصمیم‌گیری صحیح پزشکان ایجاد شده است.

علاوه بر سامانه‌هایی که به تصمیم‌گیری پزشکان کمک می‌نماید، سامانه‌های دیگری نیز وجود دارند که به خود بیمار کمک می‌نمایند. این سامانه‌ها به دنبال یافتن بهینه‌ترین روش درمانی نیستند و هدف آنها ایجاد بصیرت در فرآیند تصمیم‌گیری می‌باشد. این مهم از طریق آموزش بیمار و گوشزد کردن پیامدهای بیماری به همراه ارائه گزینه‌های پیش‌رو صورت می‌پذیرد. همچنین با کمک روان‌سنجی بیمار به صورت تلویحی اثر ترجیحات وی در درمان را به او نمایش می‌دهد.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

هر چند میان این دو نوع سیستم (مخصوص پزشکی و مخصوص بیمار) شباهت‌های زیادی وجود دارد، اما تفاوت‌های مشخصی نیز میان آنها وجود دارد. سامانه‌های مخصوص بیماران می‌بایست با حفظ دقت، تعداد سوالات از کاربر را کاهش دهد و آنها را درگیر مسائل پیچیده و تخصصی نکند.

هدف این مقاله ایجاد یک سیستم پشتیبان تصمیم‌گیری برای استخراج اطلاعات از علائم بیمار به هنگام مراجعات منظم به مرکز درمانی می‌باشد. علاوه بر استخراج این اطلاعات، می‌بایست این موارد به صورت مناسب نمایش داده شده و اقدامات پیشنهادی برای واکنش لازم ارائه گردد.

۱/۲/۲ دیابت و مالاریا

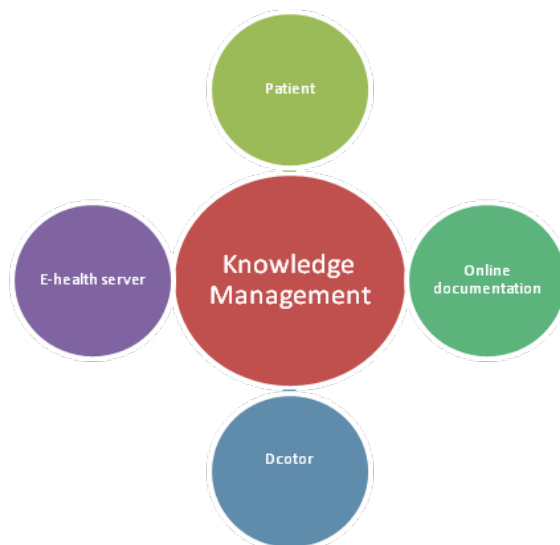
دیابت یک بیماری مزمن است که نیاز به آموزش بیمار برای خودمراقبتی دارد تا از عوارض ممکن جلوگیری شده و از ریسک وقوع عوارض بلند مدت را کاهش دهد. درمان دیابت بر کنترل میزان قند خون بیمار تمرکز دارد. این عمل به کمک دارو، رژیم غذایی و تمرینات ورزشی صورت می‌پذیرد. دیابت به دو گونه تقسیم می‌گردد. دیابت نوع اول به صورت نرمال در کودکان و نوجوانان وجود دارد و دیابت نوع دوم که شایع‌ترین نوع دیابت می‌باشد و به علت کاهش انسولین خون اتفاق می‌افتد. علت وقوع دیابت هنوز کشف نگردیده است و هم به عوامل ژنتیکی و هم عوامل محیطی همچون وزن، نژاد، جنسیت، سن و ... وابسته می‌باشد.

بیماری مالاریا نیز دارای علائمی همچون تب، سردرد، درد مفاصل، ضعف عمومی، استفراغ، تعریق و ... می‌باشد.

۱/۲/۳ مدیریت دانش

مدیریت دانش مؤثر برای ایجاد مزیت رقابتی و بهره‌وری بیشتر الزامی می‌باشد. در عصر جدید، رقابت میان سازمان‌ها تا حد زیادی به بهره‌گیری و نگهداری آنها از دانش مربوط می‌باشد. امروزه اطلاعات از منابع مختلفی همچون روزنامه‌ها، تلویزیون، اینترنت و ... قابل دسترسی می‌باشد. اما به علت عدم وجود ساختار در اطلاعات به خصوص اطلاعات اینترنتی، امکان دریافت اطلاعات غلط و نامرتب افزایش یافته است. این امر باعث ایجاد پیچیدگی در امر مدیریت دانش گردیده است. در شکل زیر مدیریت دانش در یک سیستم حفاظت از سلامت نشان داده شده است.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--



شکل ۲-۱ مدیریت دانش در سیستم حفاظت از سلامت

۱/۲/۴ مهندسی دانش

پایگاه داده‌های رابطه‌ای برای مدل سازی اطلاعات در سیستم‌های اطلاعاتی بسیار مناسب می‌باشند. سیستم‌های اطلاعاتی به منظور ثبت و پردازش اطلاعات به منظور دستیابی به یک هدف مشخص طراحی شده‌اند. پردازش داده‌ها باعث تولید اطلاعات و در ادامه تغییر اطلاعات به دانش و هوشمندی می‌شود. با شکل فعلی داده‌های سیستم‌های اطلاعاتی، امکان تولید دانش و هوشمندی از روی اطلاعات وجود ندارد. با توجه به این مطلب و برای رفع این مشکل، حوزه‌ی دیگر تحت عنوان مهندسی دانش ایجاد گردیده است. مهندسی دانش مجموعه‌ای از قواعد می‌باشد که در آن به کمک سیستم‌های رایانه‌ای به حل مسائل پیچیده‌ای که عموماً نیازمند تخصص زیادی است پرداخته می‌شود.

استخراج هستی‌شناسی یکی از مؤلفه‌های کلیدی این معماری اطلاعاتی می‌باشد. زبان هستی‌شناختی دارای معنای غیرمبهم بوده و امکان استفاده از علت‌یابی خودکار را برای طراح ایجاد می‌نماید.

۱/۲/۵ جمع‌بندی

در این مقاله یک سامانه پشتیبان تصمیم‌گیری بالینی مبتنی بر علائم برای تشخیص بیماری‌های استوایی توسعه داده شده است. رویکرد دانش‌بنیان استفاده شده در این تحقیق یک تکنیک نرم‌افزاری قدرتمند برای مشکل عدم قطعیت در داده‌های دارویی می‌باشد. این سیستم با آموزش دادن به بیمار در رابطه با بیماری، عوارض و گزینه‌های پیش‌رو برای درمان، به هدف خود یعنی یافتن بهترین روش درمان دست می‌یابد.

۱/۳ روش‌ها و ابزارهای تحلیل داده‌های حجیم

در این بخش ابزارهای و روش‌های تحلیل داده‌های حجیم مورد بررسی قرار می‌گیرد.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

۱/۳/۱ روش شماره ۱: انبارداده^۱ و مفاهیم مرتبط [3]

انبارداده به مجموعه‌ای از داده‌ها گفته می‌شود که از منابع مختلف اطلاعاتی سازمان جمع‌آوری، دسته‌بندی و ذخیره می‌شود.

انبارداده، پایگاه داده‌ای است که برای گزارش‌گیری و تحلیل داده به کار می‌رود و به عنوان هسته اصلی یک سیستم هوش تجاری^۲ یا پشتیبان تصمیم‌گیری^۳ به شمار می‌آید. به عبارت دیگر انبارداده یک مخزن داده مرکزی از داده‌های تجمیع شده است که از سیستم‌ها و منابع مختلف سازمان جمع‌آوری شده است.

انبارداده یک بانک اطلاعاتی رابطه‌ای غیرنرمال است که داده‌های حال و گذشته را در یک مکان واحد تجمیع کرده است و هدف اصلی آن پوشش گزارش‌گیری و نیازهای تحلیلی یک سازمان به کار گرفته می‌شود.

از اواسط سالهای ۱۹۸۰ نیاز به انبار داده‌ها به وجود آمد و در کنار سیستم‌های پردازش تراکنش برخط^۴ و سیستم‌های تحلیلی برخط^۵ ایجاد شد.

۱/۳/۱/۱ سیستم‌های پردازش تراکنش برخط (OLTP)

سیستم‌های OLTP معمولاً دارای ویژگی‌های زیر هستند:

- ✓ محیطی برای ورود داده‌ها و فعالیت‌های روزانه کسب‌وکار هستند.
- ✓ به دلیل کوتاهی عملیات حذف، تغییر و ویرایش اطلاعات دارای سرعت قابل توجهی می‌باشند.
- ✓ کاربران این سیستم‌ها، عموماً اپراتورها هستند.
- ✓ الگو و ساختار پایگاه داده این سیستم‌ها از فرم سوم نرمال‌سازی استفاده می‌کند.

۱/۳/۱/۲ سیستم‌های تحلیلی برخط (OLAP)

این سیستم‌ها شامل تاریخچه اطلاعات به منظور ایجاد گزارش‌های تحلیلی است تا در اختیار مدیران سازمان قرار گیرد.

- ✓ به دلیل تحلیل حجم بالایی از داده‌ها معمولاً کندتر از سیستم‌های عملیاتی می‌باشند.

¹ Data Warehouse

² Business Intelligence (BI)

³ Decision Support System (DSS)

⁴ Online Transaction Processing (OLTP)

⁵ Online Analytical Processing (OLAP)

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

- ✓ استفاده کنندگان این سیستم‌ها معمولاً مدیران و تصمیم‌گیرندگان سازمان می‌باشند.
- ✓ صرف نظر از منابع داده یک مدل داده مشترک برای تمام داده‌های مورد علاقه ارائه می‌دهد.
- ✓ ساختار پایگاه داده این سیستم‌ها معمولاً بانک اطلاعاتی رابطه‌ای غیرنرمال است.

۱/۳/۱۳ و ویژگی‌های داده‌های درون انبار داده [3]

داده‌های انبار داده از نگاه اینمون^۱ دارای ۴ ویژگی اصلی زیر هستند:

۱. غیرفرار و ماندگار (Non-Volatile)

هیچ رکوردی یا داده‌ای بروزرسانی نخواهد شد و صرفاً رکوردهایی که محتوی مقادیر جدید داده‌ها هستند، به سیستم اضافه خواهند شد.

۲. موضوع گرا (Subject-Oriented)

منظور از «موضوع» پایه‌های اساسی یک کسب و کار است، به شکلی که با حذف یکی از این پایه‌ها، شاید ماهیت آن کسب و کار تغییر یابد. به بیان دیگر هر انبار داده، داده‌های مرتبط با یک موضوع خاص را نگهداری می‌کند که این داده‌ها جهت استخراج تحلیل‌های خاص به کار گرفته می‌شوند.

۳. یکپارچه (Integrated)

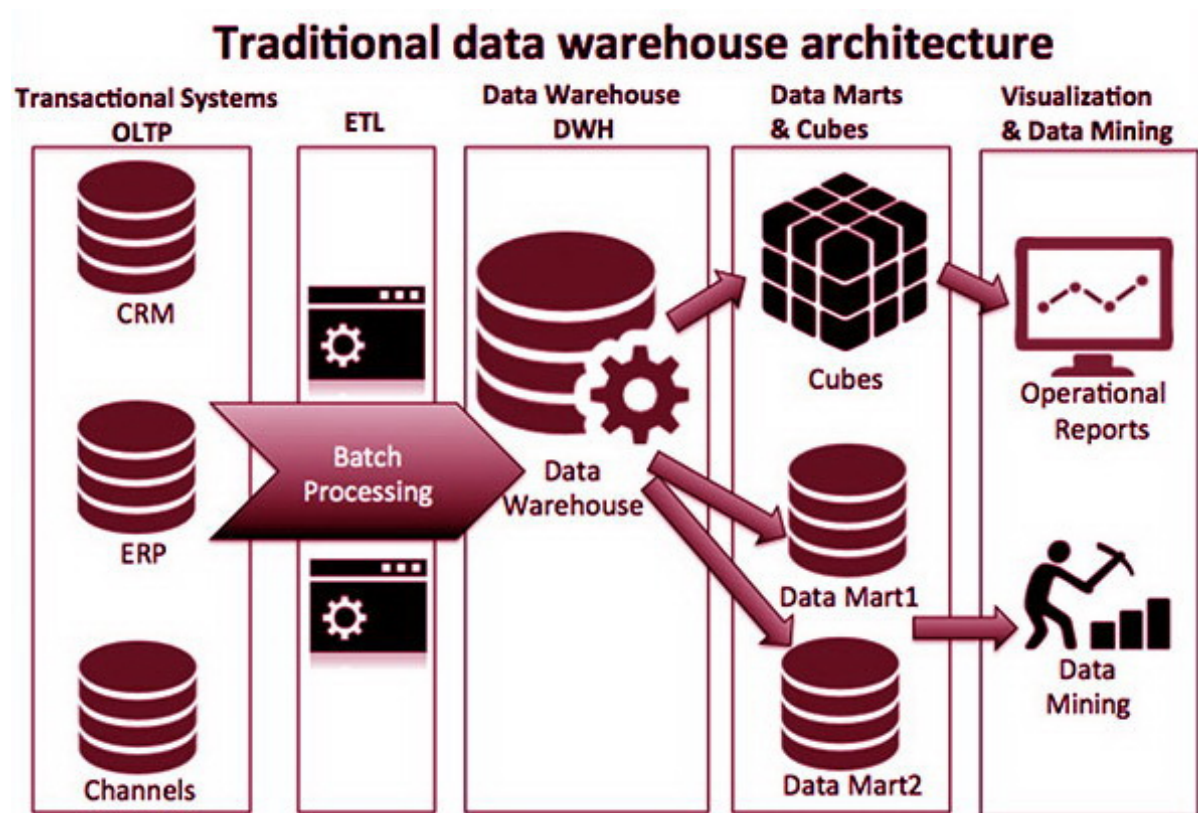
باید تمامی اطلاعاتی که در سیستم‌های عملیاتی وجود دارند و معانی یکسانی دارند، به یک روش ذخیره و نمایش داده شوند. با توجه به اینکه داده‌ها از سیستم‌ها و منابع مختلف جمع‌آوری می‌گردند تکنیک‌های مختلف یکپارچه‌سازی و تجمیع به منظور تامین یکپارچگی داده به کار گرفته می‌شود. برای نمونه، داده‌های مربوط به «جنسیت افراد» در بخشی از انبار داده به صورت «آقا/خانم» و در جای دیگری به صورت «مرد/زن» ذخیره نشده باشد.

۴. متغیر با زمان (Time Variant)

هر رکورد باید حاوی فیلد و یا کلیدی باشد که نمایانگر این باشد که این رکورد در چه زمانی ایجاد، استخراج و ذخیره شده است. از آنجا که داده‌های درون سیستم‌های عملیاتی آخرین و به‌روزترین داده هر سیستم می‌باشد، نیازی به وجود چنین عنصری در سیستم‌های OLTP احساس نمی‌گردد، ولی چون در انبار داده تمام داده‌های نسخ قدیمی داده‌های سیستم‌های عملیاتی موجود می‌باشد، باید حتماً مشخص گردد که هر داده‌ای در سیستم‌های عملیاتی در چه زمانی، چه مقادیری داشته است.

¹ Bill Inmon

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--



شکل ۳-۱ مدل سنتی انبار داده

۱/۳/۲ ابزار شماره ۲: Apache Cassandra [4]

کاساندرایک پایگاه داده غیر رابطه ای^۱ و توزیع شده است. این پایگاه داده برای مدیریت داده های عظیم (Big Data) کاربرد فراوانی دارد. کاساندرایک هیچ نقطه خاص خرابی^۲ ندارد و به راحتی می تواند بر روی چندین کامپیوتر توزیع شود. این پایگاه داده توسعه پذیر خطی^۳ است و دسترس پذیری^۴ بالایی دارد. کاساندرایک برای کار با داده های بسیار بزرگ طراحی شده است تا کاربران بتوانند این داده ها را به راحتی و با سرعت بالا دریافت کنند. کاساندرایک از زبان پرس و جویی به نام CQL استفاده می کند که بسیار شبیه SQL است و برنامه نویسانی که می خواهند از SQL به کاساندرایک مهاجرت کنند، کار زیاد سختی در پیش ندارند.

¹ NoSQL

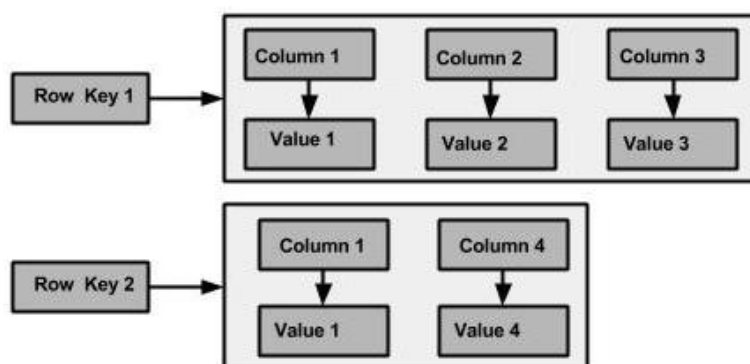
² Single Point Of Failure

³ Linear Scalable

⁴ Availability

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

همانطور که اشاره گردید، کاساندرا از دسته پایگاه داده‌های غیر رابطه‌ای و از نوع ستونی^۱ است. این پایگاه داده‌ها که داده‌ها را به جای اینکه به صورت سطر به سطر مشاهده کنند، به صورت ستون به ستون مشاهده و ذخیره سازی می کنند، مناسب برای کاربردهایی مانند پردازش داده‌های عظیم هستند. در کاساندرا دیگر پیش شرطی برای یکسان بودن تعداد و نوع ستون‌های یک جدول وجود ندارد و سطرهای مختلف می‌توانند دارای ستون‌های متفاوتی باشند.



شکل ۴-۱ شمای داده‌ها در پایگاه داده‌ی آپاچی کاساندرا

در کاساندرا از معماری Master/Slave استفاده نشده است. زیرا در این معماری معمولاً گره سرپرست^۲ به دلیل کارکرد زیاد، به گلوگاه^۳ سیستم تبدیل می‌شود. یعنی در صورتی که سرپرست از کار بیافتد یا کند شود، کل سیستم از کار می‌افتد یا کند می‌شود. برای همین در کاساندرا چیزی به نام Master وجود ندارد و تمامی گره‌های متصل به هم، مانند یکدیگر رفتار می‌کنند و داده‌ها در گره‌های مختلف تکثیر^۴ می‌شوند. ساختار تکرار و تکثیر در آپاچی کاساندرا مانند شکل زیر است:

^۱ Column Family Database

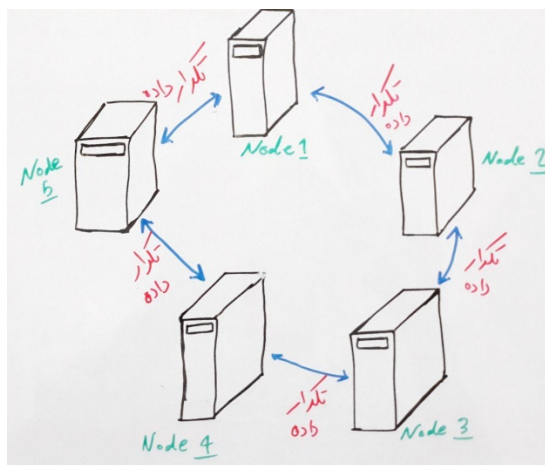
^۲ Master

^۳ Bottleneck

^۴ Node

^۵ Replicate

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--



شکل ۵-۱ معماری کلان پایگاه داده‌ی آپاچی کاساندر

همان طور که در شکل بالا مشاهده می‌گردد، داده‌ها به صورت نظیر به نظیر^۱ بر روی گره‌های مختلف تکثیر می‌شوند.

۱/۳/۳ ابزار شماره ۳: کتابخانه fastText [5]

کتابخانه fastText یک کتابخانه به زبان پایتون برای دسته‌بندی و نمایش متون می‌باشد. این کتابخانه متن‌باز بوده، رایگان و بسیار سبک می‌باشد. fastText توسط بنیاد پروژه‌های متن‌باز شرکت Facebook پشتیبانی می‌شود.

❖ دسته‌بندی متن^۲: هدف از دسته‌بندی متن، تقسیم‌بندی اسنادی همچون ایمیل، بلاگ، پیغام‌های متنی، بازخورد مشتریان و ... به یک یا چند دسته می‌باشد. نمونه‌هایی از این دسته‌بندی می‌تواند به صورت زیر باشد:

- ✓ ایمیل هرز^۳ و غیر هرز
- ✓ دسته‌بندی نظرات مشتریان بر اساس امتیازی که به محصول داده‌اند
- ✓ دسته‌بندی اسناد بر اساس زبان آنها
- ✓ و ...

امروزه روش منتخب برای انجام اینکار، یادگیری ماشین است که قوانین دسته‌بندی را از روی مثال‌ها به دست می‌آورد.

¹ Peer to peer (P2P)

² Text classification

³ Spam

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

❖ نمایش لغت^۱ (با بردار): یک ایده مدرن یادگیری ماشین نمایش لغات با استفاده از بردار است. این بردارها معانی مخفی از اطلاعات مربوط به زبان متن ورودی همچون معنا و شباهت‌ها به دست می‌آورد. برای واضح‌شدن موضوع یک نمونه فرضی در ادامه ارائه شده است:

	King	Queen	Woman	Princess	...
Royalty	0.99	0.99	0.02	0.98	
Masculinity	0.99	0.05	0.01	0.02	
Femininity	0.05	0.93	0.999	0.96	
Age	0.7	0.6	0.5	0.1	
...					

شکل 7-1 نمونه‌ای فرضی از نمایش لغت با بردار

دسته‌بندی متن و نمایش لغت هر دو از امکاناتی هستند که توسط کتابخانه fastText قابل انجام می‌باشند.

۱/۳/۴ ابزار شماره ۴: NLTK^۲ [6]

NLTK یک بستر بسیار قدرتمند در پایتون برای پردازش زبان طبیعی می‌باشد. NLTK یک بستر رایج برای ساخت برنامه‌هایی به منظور تجزیه و تحلیل متون مختلف است.

به پردازش زبان انسان به صورت خودکار یا نیمه خودکار، پردازش زبان طبیعی^۳ گفته می‌شود. NLP دارای طیف گسترده‌ای از برنامه‌های کاربردی در زمینه‌های مختلفی مانند بهداشت و سلامت، آموزش، تجارت و ... است. پردازش زبان طبیعی در علوم کامپیوتر به حوزه‌هایی مانند نظریه زبان‌ها، تکنیک‌های کامپایلر، تعامل انسان با کامپیوتر و یادگیری ماشین بسیار وابسته است.

برای کار با NLTK لازم است تا در ابتدا مجموعه‌ای از متون را بارگذاری شود. این مجموعه متون که با نام corpus نیز شناخته می‌شوند، از طریق NLTK قابل دستیابی هستند. یک corpus که صورت جمع آن corpora است، در Wikipedia به صورت زیر تعریف می‌شود:

«به مجموعه‌ای خام از داده‌های زبانی نوشتاری یا گفتاری گفته می‌شود که می‌توان در توصیف و تحلیل زبان از آن بهره گرفت.»

¹ Word representations

² Natural Language Toolkit

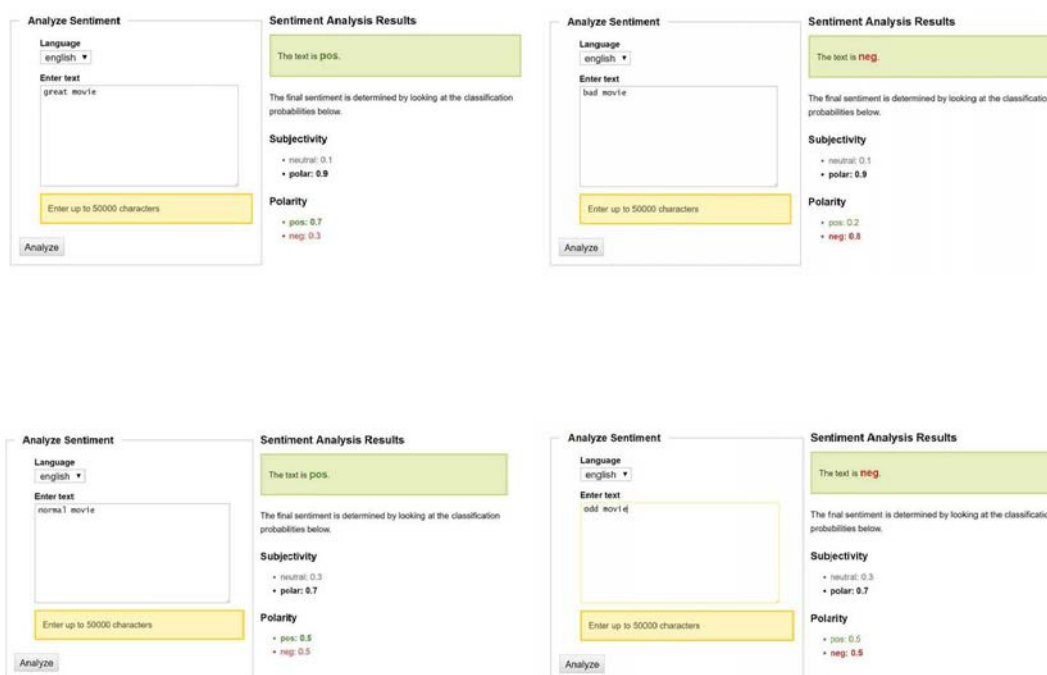
³ Natural-language processing (NLP)

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

بنابراین می‌توان گفت یک corpus، حجم وسیعی از فایل‌های متنی را شامل می‌شود. NLKT دارای مجموعه‌ای از corpus های آماده می‌باشد که امر تجزیه و تحلیل متون را بسیار آسان می‌نماید.

باید توجه داشت که از آن جا که پردازش زبان طبیعی بر دو جنبه‌ی متن و صوت زبان کار می‌کند، بنابراین یک corpus به صورت مجموعه‌ای از فایل‌های صوتی نیز می‌تواند وجود داشته باشد. اما در این متن تنها بر جنبه‌ی متنی آن تمرکز گردیده است.

یکی از مهمترین تحلیل‌های که به کمک NLKT می‌توان انجام داد، تحلیل احساس^۱ متن می‌باشد. این تحلیل بسیار کاربردی بوده و در بررسی دیدگاه‌ها و نظرات مشتریان یک محصول یا خدمت بسیار مفید می‌باشد. برای نمونه در تصویر زیر چند نمونه تحلیل بر روی جملات ساده صورت پذیرفته است:



The figure displays four screenshots of the NLTK Sentiment Analysis web interface, arranged in a 2x2 grid. Each screenshot shows the 'Analyze Sentiment' form on the left and the 'Sentiment Analysis Results' on the right.

- Top Left:** Input text is 'great movie'. The result shows 'The text is pos.' (positive) with Subjectivity (neutral: 0.1, polar: 0.9) and Polarity (pos: 0.7, neg: 0.3).
- Top Right:** Input text is 'bad movie'. The result shows 'The text is neg.' (negative) with Subjectivity (neutral: 0.1, polar: 0.9) and Polarity (pos: 0.2, neg: 0.8).
- Bottom Left:** Input text is 'normal movie'. The result shows 'The text is pos.' (positive) with Subjectivity (neutral: 0.3, polar: 0.7) and Polarity (pos: 0.5, neg: 0.5).
- Bottom Right:** Input text is 'odd movie'. The result shows 'The text is neg.' (negative) with Subjectivity (neutral: 0.3, polar: 0.7) and Polarity (pos: 0.5, neg: 0.5).

شکل ۷-۱ تحلیل احساس متن به کمک NLTP

¹ Sentiment Analysis

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

۱/۳/۵ روش شماره ۵: رگرسیون خطی [7]

در مدل‌های آماری، تحلیل رگرسیون یک فرایند آماری برای تخمین روابط بین متغیرها می‌باشد. این روش شامل تکنیک‌های زیادی برای مدل‌سازی و تحلیل متغیرهای خاص و منحصر بفرد، با تمرکز بر رابطه بین متغیر وابسته و یک یا چند متغیر مستقل، می‌باشد. تحلیل رگرسیون خصوصاً کمک می‌کند در فهم اینکه چگونه مقدار متغیر وابسته با تغییر هر کدام از متغیرهای مستقل و با ثابت بودن دیگر متغیرهای مستقل تغییر می‌کند. در همه موارد هدف تخمین یک تابع از متغیرهای مستقل است که تابع رگرسیون نامیده شده است.

یکی از انواع تحلیل رگرسیون، نوع خطی آن می‌باشد که به کمک این تحلیل می‌توان براساس داده‌های گذشته یک تابع رگرسیون خطی برای مسئله یافته و به کمک آن داده‌های ناموجود را که معمولاً داده‌های زمان آتی می‌باشند، محاسبه نمود.

برای نمونه در ساده‌ترین حالت ممکن به کمک N داده‌ی موجود و یک متغیر مستقل (X_i) خواهیم داشت:

$$Y_i = \beta_0 + \beta_1 X_i$$

به کمک روش کمترین مربعات (محاسبه مجموع مربعات خطا و حداقل کردن آن به کمک مشتق‌گیری) ضرایب β_0 و β_1 به صورت زیر خواهند بود:

$$\beta_1 = \frac{\sum (\bar{x} - x_i)(\bar{y} - y_i)}{\sum (\bar{x} - x_i)^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

۱/۳/۶ روش شماره ۶: الگوریتم C4.5 [8]

الگوریتم C4.5 یک الگوریتم قدرتمند برای ایجاد درخت تصمیم به منظور دسته‌بندی اطلاعات می‌باشد. در حقیقت این الگوریتم به کمک یک سری ورودی دسته‌بندی شده، یک درخت تصمیم ایجاد کرده و بر اساس آن می‌تواند داده‌های جدید را دسته‌بندی نماید. این الگوریتم توسط راس کوینلن^۱ و پس از الگوریتم ID3 طراحی گردید. راس کوینلن پس از اینکه به نقاط ضعف الگوریتم ID3 پی برد، در مدت کوتاهی الگوریتم بعدی خود یعنی C4.5 را طراحی کرد.

ویژگی‌ها و برتری‌های این الگوریتم عبارتند از:

۱. الگوریتم C4.5 می‌تواند مقادیر گسسته یا پیوسته را در ویژگی‌ها درک کند.

¹ Ross Quinlan

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

۲. این الگوریتم قادر است تا مقادیری که موجود نیستند را هم تحمل کند.
۳. سومین موردی که باعث بهینه شدن الگوریتم C4.5 می شود، عملیات هرس کردن^۱ جهت جلوگیری از Overfitting می باشد. الگوریتم‌هایی مانند ID3 به دلیل اینکه سعی دارند تا حد امکان شاخه و برگ داشته باشند (تا به نتیجه مورد نظر برسند) با احتمال بالاتری دارای پیچیدگی در ساخت مدل می باشند و این پیچیدگی در بسیاری از موارد الگوریتم را دچار Overfitting و خطای بالا می کند. اما با عملیات هرس کردن درخت که در الگوریتم C4.5 نیز انجام می شود، می توان مدل را به یک نقطه بهینه رساند که زیاد پیچیده نباشد و Overfitting رخ ندهد.
۴. ویژگی دیگر الگوریتم C4.5 که می تواند این الگوریتم را از سایر الگوریتم‌های مشابه، متمایز کند، قابلیت وزن دهی به شاخص‌ها است.
- برای نمونه مسئله روبه‌رو را در نظر بگیرید: «چه افرادی در آزمون دکترا در دانشگاه پذیرفته می شوند؟»
- شاخص‌های زیر برای قبولی فرد در آزمون دکترا در نظر گرفته شده است:

- معدل کل
- تعداد مقالات
- مدرک زبان
- سنوات تحصیلی

یک مجموعه داده‌ی ورودی نیز برای الگوریتم C4.5 می‌بایست فراهم گردد:

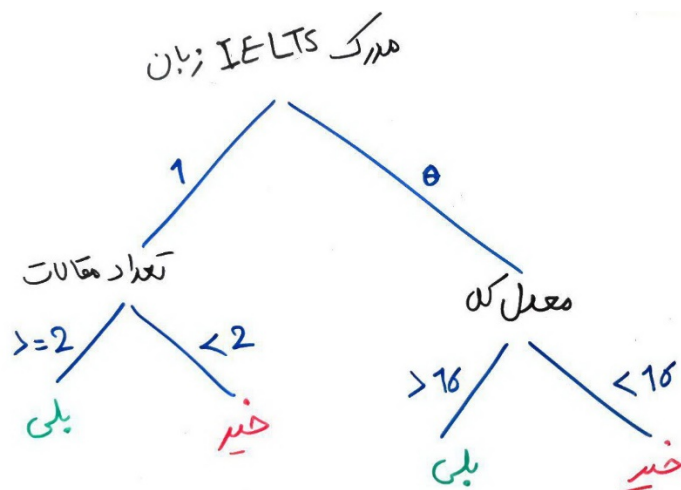
¹ Pruning

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

ردیف	معدل کل	تعداد مقالات	مدرک زبان IELTS	سنوات تحصیلی	دکتری قبول شده؟
#1	19,5	3	1	3	بلی
#2	16,5	0	1	4	خیر
#3	15	0	0	3	خیر
#4	17	2	1	2,5	بلی
#5	18,5	2	0	2,5	بلی
#6	15,5	1	1	2,5	خیر
#7	19	3	1	3	بلی

شکل ۸-۱ داده‌های ورودی برای الگوریتم C4.5

نمونه خروجی درخت تصمیم با داده‌های ارائه شده، به صورت زیر خواهد بود که بر اساس آن می‌توان در مورد افراد جدید و قبولی آنها در آزمون دکترا قضاوت انجام داد.



شکل ۹-۱ درخت تصمیم خروجی از الگوریتم C4.5

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

گام دوم

در این گام با بهره‌گیری از روش‌ها و ابزارهای گام نخست، چند مورد از نتایجی را که می‌توان از داده‌های شرکت استخراج نمود، ارائه می‌نماییم:

گزارش شماره ۱	
توضیح اولیه گزارش	لیستی از ۲۰ توییت‌کننده مؤثر جهانی
نتایج استخراج‌شده	در این گزارش نام کاربری، ۲۰ مورد از تأثیرگذارترین کاربران شبکه اجتماعی توییتر به همراه امتیاز اثرگذاری هر کدام از این افراد استخراج می‌گردد.
موارد استفاده از نتایج	می‌توان میان افراد تأثیرگذار و شرکت (بخش طراحی محصول، بخش بازاریابی و تبلیغات و ...) ارتباط برقرار کرده و علاوه بر استفاده از نظرات ایشان، از پتانسیل این افراد برای جهت‌دهی به افکار عمومی استفاده نمود.
داده‌های مورد نیاز	توییت‌ها، تعداد پسندیده‌شدن هر توییت، تعداد بازتوییت شدن هر توییت، نام کاربری توییت‌کننده به همراه تعداد افراد دنبال‌کننده و دنبال‌شونده
ابزارها و روش‌ها	بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل توییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود. روش مورد استفاده به صورت نسبت دادن امتیاز به هر کاربر، بر اساس تعداد توییت، بازتوییت، پسندها، دنبال‌شونده‌ها و دنبال‌کننده‌ها می‌باشد.
سطح سختی استخراج گزارش و قابلیت خودکارسازی	آسان - قابل خودکارسازی ^۱
سطح DIKW	نتایج استخراج‌شده در سطح دانش قرار می‌گیرند. ^۲

^۱ به طور کلی هر گزارشی قابل خودکارسازی است و امکان خودکارسازی به زیرساخت‌ها، تکنولوژی مورد استفاده برای ذخیره‌بازیابی داده بستگی دارد.

^۲ بر اساس ضمیمه شماره ۱

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

گزارش شماره ۲	
لیستی از ۱۰ توییت کننده مؤثر جهانی به تفکیک کشور	توضیح اولیه گزارش
در این گزارش نام کاربری، ۱۰ مورد از تأثیرگذارترین کاربران شبکه اجتماعی توییتر به همراه امتیاز اثرگذاری هر کدام از این افراد به تفکیک کشور استخراج می‌گردد.	نتایج استخراج شده
می‌توان میان افراد تأثیرگذار در هر کشور و شعبه مربوطه شرکت (بخش طراحی محصول، بخش بازاریابی و تبلیغات و ...) ارتباط برقرار کرده و علاوه بر استفاده از نظرات ایشان، از پتانسیل این افراد برای جهت‌دهی به افکار عمومی استفاده نمود.	موارد استفاده از نتایج
توییت‌ها، تعداد پس‌نویس‌شدن هر توییت، تعداد بازتوییت شدن هر توییت، نام کاربری توییت کننده به همراه تعداد افراد دنبال کننده و دنبال‌شونده، محل جغرافیایی ثبت هر توییت (به کمک ip و یا محل زندگی ارائه شده توسط خود کاربران توییتر)	داده‌های مورد نیاز
بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل توییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها و روش‌ها
روش مورد استفاده به صورت نسبت دادن امتیاز به هر کاربر، بر اساس تعداد توییت، بازتوییت، پس‌نویس‌ها، دنبال‌شونده‌ها و دنبال‌کننده‌ها می‌باشد.	
آسان - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح دانش قرار می‌گیرند.	سطح DIKW

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۳	
توضیح اولیه گزارش	ساعاتی از روز که بیشترین فعالیت در توییتر وجود دارد
نتایج استخراج شده	در این گزارش نموداری از ساعات روز به همراه تعداد توییت های انتشار یافته در آن ساعت ارائه می شود و ۳ ساعتی که بیشترین و کمترین فعالیت در آن وجود دارد مشخص می گردد.
موارد استفاده از نتایج	به کمک دانستن زمان های اصلی فعالیت در توییتر، بخش اطلاع رسانی، بازاریابی و تبلیغات شرکت می تواند توییت های مهم خود را در این ساعات از روز انتشار دهد. همچنین خبرهای کم اهمیت و خبرهایی که ترجیح داده می شود مسکوت باقی بمانند در ساعات خلوت منتشر گردند.
داده های مورد نیاز	توییت ها، زمان نشر توییت
ابزارها و روش ها	بسته به نوع ذخیره داده ها از ابزارهای متنوعی می توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه های تحلیل توییت می توانند برای این گزارش مورد استفاده قرار گیرند. داده های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می بایست از ابزارهایی همچون pivot table استفاده نمود.
سطح سختی استخراج گزارش و قابلیت خودکارسازی	آسان - قابل خودکارسازی
سطح DIKW	نتایج استخراج شده در سطح دانش قرار می گیرند.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۴	
توضیح اولیه گزارش	ساعاتی از روز که بیشترین فعالیت در توییتر وجود دارد به تفکیک کشور
نتایج استخراج شده	در این گزارش نموداری از ساعات روز به همراه تعداد توییت های انتشار یافته در آن ساعت ارائه می شود و ۳ ساعتی که بیشترین و کمترین فعالیت در آن وجود دارد به تفکیک کشور مشخص می گردد.
موارد استفاده از نتایج	به کمک دانستن زمان های اصلی فعالیت در توییتر به تفکیک کشور، بخش اطلاع رسانی، بازاریابی و تبلیغات شرکت می تواند توییت های مهم خود را در این ساعات از روز انتشار دهد. همچنین خبرهای کم اهمیت و خبرهایی که ترجیح داده می شود مسکوت باقی بمانند در ساعات خلوت منتشر گردند. با توجه به تفاوت ساعت کار، خواب و ... در کشورهای مختلف برای تبلیغات و اطلاع رسانی محلی می توان از نتایج حاصله بهره برد.
داده های مورد نیاز	توییت ها، زمان نشر توییت، محل جغرافیایی ثبت هر توییت (به کمک ip و یا محل زندگی ارائه شده توسط خود کاربران توییتر)
ابزارها و روش ها	بسته به نوع ذخیره داده ها از ابزارهای متنوعی می توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه های تحلیل توییت می توانند برای این گزارش مورد استفاده قرار گیرند. داده های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می بایست از ابزارهایی همچون pivot table استفاده نمود.
سطح سختی استخراج گزارش و قابلیت خودکارسازی	آسان - قابل خودکارسازی
سطح DIKW	نتایج استخراج شده در سطح دانش قرار می گیرند.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۵	
میزان مثبت یا منفی بودن (بار احساسی) توییت های منتشر شده	توضیح اولیه گزارش
در این گزارش میزان مثبت یا منفی بودن توییت های منتشر شده به صورت یک عدد نمایش داده می شود	نتایج استخراج شده
نتیجه این گزارش می تواند به عنوان یک باز خورد کلی از وضعیت محبوبیت و رضایت کاربران از فعالیتهای شرکت در نظر گرفته شود. این گزارش باید به صورت بلادرنگ ^۱ تولید شده و یک بازه ی نو سان قابل قبول برای آن در نظر گرفته شود. در صورتی که نتیجه گزارش از بازه ی تعریف شده خارج شود، علت این مسئله می بایست پیدا شده و مسئله بررسی و رفع گردد.	موارد استفاده از نتایج
محتوای توییت ها، تعداد پسندیده شدن هر توییت، تعداد بازتوییت شدن هر توییت	داده های مورد نیاز
بسته به نوع ذخیره داده ها از ابزارهای متنوعی می توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه های تحلیل توییت می توانند برای این گزارش مورد استفاده قرار گیرند. داده های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
علاوه بر این می بایست بر روی متن توییت ها تحلیل احساسات صورت گرفته و میزان مثبت یا منفی بودن آن معین گردد.	
سخت - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح دانش قرار می گیرند.	سطح DIKW

¹ Real-time

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۶	
میزان مثبت یا منفی بودن (بار احساسی) توییت های منتشر شده به تفکیک کشور	توضیح اولیه گزارش
در این گزارش به ازای هر کشور میزان مثبت یا منفی بودن توییت های منتشر شده به صورت یک عدد نمایش داده می شود	نتایج استخراج شده
نتیجه این گزارش می تواند به عنوان یک باز خورد کلی از وضعیت محبوبیت و رضایت کاربران از فعالیتهای شرکت در یک کشور خاص در نظر گرفته شود. این گزارش باید به صورت بلادرنگ تولید شده و یک بازه ی نو سان قابل قبول برای آن در نظر گرفته شود. در صورتی که نتیجه گزارش از بازه ی تعریف شده خارج شود، علت این مسئله می بایست پیدا شده و مسئله بررسی و رفع گردد.	موارد استفاده از نتایج
محتوای توییت ها، تعداد پسندیده شدن هر توییت، تعداد بازتوییت شدن هر توییت	داده های مورد نیاز
بسته به نوع ذخیره داده ها از ابزارهای متنوعی می توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه های تحلیل توییت می توانند برای این گزارش مورد استفاده قرار گیرند. داده های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
علاوه بر این می بایست بر روی متن توییت ها تحلیل احساسات صورت گرفته و میزان مثبت یا منفی بودن آن معین گردد.	
سخت - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح دانش قرار می گیرند.	سطح DIKW

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

گزارش شماره ۷	
میزان مثبت یا منفی بودن (بار احساسی) توییت های منتشر شده به تفکیک محصول	توضیح اولیه گزارش
در این گزارش به ازای هر محصول میزان مثبت یا منفی بودن توییت های منتشر شده به صورت یک عدد نمایش داده می شود	نتایج استخراج شده
نتیجه این گزارش می تواند به عنوان یک باز خورد کلی از وضعیت محبوبیت و رضایت کاربران از محصولات شرکت در نظر گرفته شود. این گزارش باید به صورت بلادرنگ تولید شده و یک بازهی نوسان قابل قبول برای آن در نظر گرفته شود. در صورتی که نتیجه گزارش از بازهی تعریف شده خارج شود، علت این مسئله می بایست پیدا شده و مسئله بررسی و رفع گردد.	موارد استفاده از نتایج
محتوای توییت ها، تعداد پسندیده شدن هر توییت، تعداد بازتوییت شدن هر توییت، محصولات شرکت	داده های مورد نیاز
بسته به نوع ذخیره داده ها از ابزارهای متنوعی می توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه های تحلیل توییت می توانند برای این گزارش مورد استفاده قرار گیرند. داده های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
علاوه بر این می بایست بر روی متن توییت ها تحلیل احساسات صورت گرفته و میزان مثبت یا منفی بودن آن معین گردد.	سطح سختی استخراج گزارش و قابلیت خودکارسازی
سخت - قابل خودکارسازی	
نتایج استخراج شده در سطح دانش قرار می گیرند.	سطح DIKW

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۸	
متن ۲۰ توییت برتر	توضیح اولیه گزارش
در این گزارش ابتدا ۲۰ توییت برتر با توجه به تعداد پسندها و بازتوییت‌ها مشخص گردیده و متن آنها برای بررسی بیشتر در گزارش ذکر می‌گردد.	نتایج استخراج شده
می‌توان با بررسی ۲۰ توییت برتر، رضایت کاربران، حوزه‌های مورد توجه کاربران، پیشنهادات و انتقادات مطرح شده و ... را مشاهده و بر اساس آن در تمام بخش‌های سازمان اقدامات بهبودی را انجام داد.	موارد استفاده از نتایج
توییت‌ها، تعداد پسندیده شدن هر توییت، تعداد بازتوییت شدن هر توییت	داده‌های مورد نیاز
بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل توییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
آسان - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح اطلاعات قرار می‌گیرند.	سطح DIKW

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۹	
۲۰ مورد از پرتکرارترین هشتگ‌ها	توضیح اولیه گزارش
در این گزارش ۲۰ هشتگ پرتکرار به همراه تعداد تکرار آنها در یک جدول به عنوان خروجی ارائه می‌گردد.	نتایج استخراج شده
می‌توان با بررسی ۲۰ هشتگ پرتکرار، رضایت کاربران، حوزه‌های مورد توجه کاربران، پیشنهادات و انتقادات مطرح شده و ... را مشاهده و بر اساس آن در تمام بخش‌های سازمان اقدامات بهبودی را انجام داد. به خصوص انتقادات کاربران که در اکثر موارد به صورت هشتگ بیان می‌گردند.	موارد استفاده از نتایج
توییت‌ها، هشتگ‌های استفاده شده	داده‌های مورد نیاز
بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل توییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
آسان - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح اطلاعات قرار می‌گیرند.	سطح DIKW

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۱۰	
توضیح اولیه گزارش	۲۰ مورد از پر تکرارترین هشتگ‌ها به تفکیک کشور
نتایج استخراج شده	در این گزارش ۲۰ هشتگ پرتکرار به تفکیک کشور به همراه تعداد تکرار آنها در یک جدول به عنوان خروجی ارائه می‌گردد.
موارد استفاده از نتایج	می‌توان با بررسی ۲۰ هشتگ پرتکرار، رضایت کاربران، حوزه‌های مورد توجه کاربران، پیشنهادات و انتقادات مطرح شده و ... را مشاهده و بر اساس آن در تمام بخش‌های سازمان اقدامات بهبودی را انجام داد. به خصوص انتقادات کاربران که در اکثر موارد به صورت هشتگ بیان می‌گردند.
داده‌های مورد نیاز	توییت‌ها، هشتگ‌های استفاده شده، محل جغرافیایی ثبت هر توییت (به کمک ip و یا محل زندگی ارائه شده توسط خود کاربران توییت‌ر)
ابزارها	بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل توییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود.
سطح سختی استخراج گزارش و قابلیت خودکارسازی	آسان - قابل خودکارسازی
سطح DIKW	نتایج استخراج شده در سطح اطلاعات قرار می‌گیرند.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

گزارش شماره ۱۱	
لیستی ۱۰ زبان پر استفاده برای توییت	توضیح اولیه گزارش
در این گزارش نموداری از ۱۰ زبان پر استفاده در توییت‌ها به همراه تعداد توییت به آن زبان ارائه می‌گردد.	نتایج استخراج شده
از نتایج این گزارش می‌توان در راستای شناسایی زبان‌هایی که می‌بایست بخش تبلیغات و اطلاع‌رسانی شرکت بر آن تمرکز نماید، استفاده نمود.	موارد استفاده از نتایج
توییت‌ها، زبان توییت	داده‌های مورد نیاز
بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل توییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
آسان - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح اطلاعات قرار می‌گیرند.	سطح DIKW

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

گزارش شماره ۱۲	
لیستی ۱۰ کشوری که بیشترین فعالیت را در توییتر دارند	توضیح اولیه گزارش
در این گزارش نموداری از ۱۰ کشور اول که بیشترین استفاده از توییتر را دارند، ارائه می‌گردد.	نتایج استخراج شده
از نتایج این گزارش می‌توان در راستای شناسایی کشورهای که می‌بایست بخش تبلیغات و اطلاع‌رسانی شرکت بر آن تمرکز نماید، استفاده نمود.	موارد استفاده از نتایج
توییتهای، محل جغرافیایی ثبت هر تویییت (به کمک ip و یا محل زندگی ارائه شده توسط خود کاربران توییتر)	داده‌های مورد نیاز
بسته به نوع ذخیره داده‌ها از ابزارهای متنوعی می‌توان بهره برد. زبان SQL، ابزارهای اکسل و سامانه‌های تحلیل تویییت می‌توانند برای این گزارش مورد استفاده قرار گیرند. داده‌های نمونه که برای تحلیل در اختیار قرار دارد به صورت فایل اکسل بوده و بنابراین می‌بایست از ابزارهایی همچون pivot table استفاده نمود.	ابزارها
آسان - قابل خودکارسازی	سطح سختی استخراج گزارش و قابلیت خودکارسازی
نتایج استخراج شده در سطح اطلاعات قرار می‌گیرند.	سطح DIKW

❖ لازم به ذکر است که تمام گزارش‌ها به صورت مستقل از زمان در نظر گرفته شده و می‌توان همین گزارشات را با در نظر گرفتن زمان به صورت مقایسه‌ای ارائه نمود. (برای مثال مقایسه هشتگ‌های پرتکرار هر ماه با ماه گذشته)

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	---

گام سوم

در این گام گزارش‌های مشخص شده در گام دوم ایجاد و ارائه گردیده است:

* تمام فایل‌ها و کدهای مربوط به گزارش‌گیری در قالب یک پروژه گیت‌هاب به آدرس زیر قابل دسترسی است:

<https://github.com/Ashafaat/DSS>

۳/۱ گزارش شماره ۱

عنوان گزارش: لیستی از ۲۰ توییت‌کننده مؤثر جهانی

رتبه توییت‌کننده	نام کاربری توییت‌کننده	امتیاز توییت‌کننده
۱	Vala Afshar	108169545351
۲	Phil Davis	81134441100
۳	CNNMoney	3129252509
۴	Wall Street Journal	1376009855
۵	BK Stocks	739731195
۶	Business Insider	720521983
۷	Reuters Top News	438612105
۸	StockTwits	405424817
۹	Stocks On High Alert	382549174
۱۰	Downtown Josh Brown	359256045
۱۱	AppleInsider	224300835
۱۲	CNBC	170861907
۱۳	MarketWatch	148548031
۱۴	Yahoo Finance	121255841
۱۵	Reuters Business	116827692
۱۶	Aswath Damodaran	113248166
۱۷	Jim Cramer	90664923
۱۸	BI Tech	78852095
۱۹	Kourt Kardashian	55275561
۲۰	CNBC Now	50085542

- فرمول محاسبه امتیاز توییت به صورت زیر می‌باشد (ابتکاری):

$$\text{Score} = (\text{Fav} \times \text{Retweet} + 1) \times (\text{Follower} \div (\text{Following} + 1) + \text{Lists})$$

- امتیاز هر کاربر از طریق مجموع امتیاز تمام توییت‌های او محاسبه گردیده است.

- جدول نهایی از طریق قابلیت Pivot Table در اکسل بدست آمده است.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	--

۳/۲ گزارش شماره ۲

عنوان گزارش: لیستی از ۱۰ توییتکننده مؤثر جهانی به تفکیک کشور

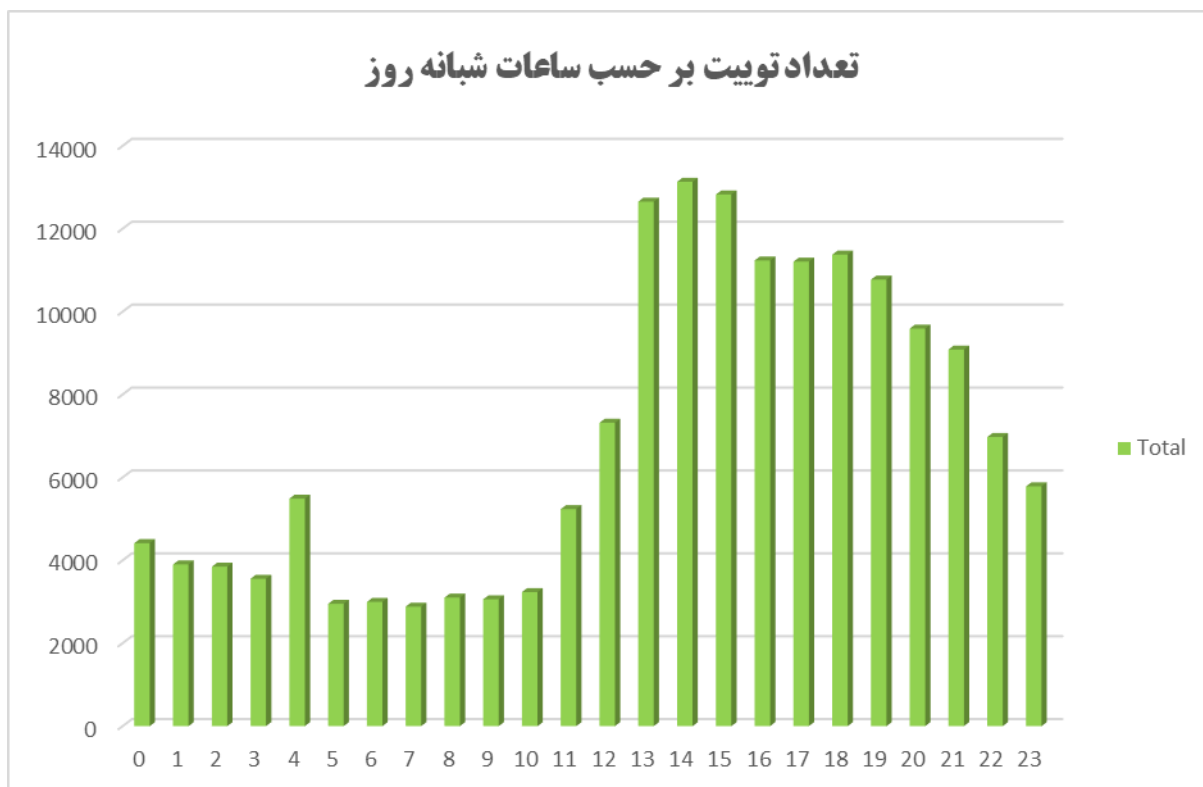
حجم این گزارش بسیار زیاد بوده و شامل حدوداً ۱۶۶ گزارش (به تعداد کشورهای موجود در توییت‌ها) می‌باشد. در این بخش تنها گزارش‌های ۳ کشور به عنوان نمونه ارائه گردیده است. (تمام گزارش‌ها در اکسل مربوطه در دسترس می‌باشد).

UA (Ukraine)		
	Username	Score
1	Sergey Kochergan	862
2	Denis Do	300
3	Michael Babich	290
4	Denys Zhadanov	256
5	Nick Hencher	93
6	casey zimmerman	57
7	megalos.austin	42
8	WhiteWolf	25
9	Danny	21
10	Roman Korobchan	14
AR (Argentina)		
	Username	Score
1	Gustavo Neffa	18879
2	R.J..I	1726
3	Fernando Camusso	977
4	Jose Prats 3T	864
5	Ricardo Sametband	446
6	Pattie Jabbaz	293
7	Nicolas Litvinoff	220
8	Mauro Cognetta	187
9	Omar (Gangi)	102
10	Alejandro Burato	100
RU (Russia)		
	Username	Score
1	TradingFloor	3054
2	James DePorre	2302
3	Eldar Murtazin	2194
4	Лёха	2045
5	Uprise	140
6	Maxim Melnikov	75
7	Dmitry Kabanov	51
8	Ирина Новикова	29
9	Kriish	28
10	Anton Lyubushkin	27

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

۳/۳ گزارش شماره ۳

عنوان گزارش: ساعاتی از روز که بیشترین فعالیت در توییت وجود دارد



- بیشترین فعالیت: ساعات ۱۳ الی ۱۶ (اوایل بعد از ظهر)
- کمترین فعالیت: ساعات ۵ الی ۸ (صبح زود)
- نمودار نهایی از طریق قابلیت Pivot Table و رسم نمودار در اکسل بدست آمده است.

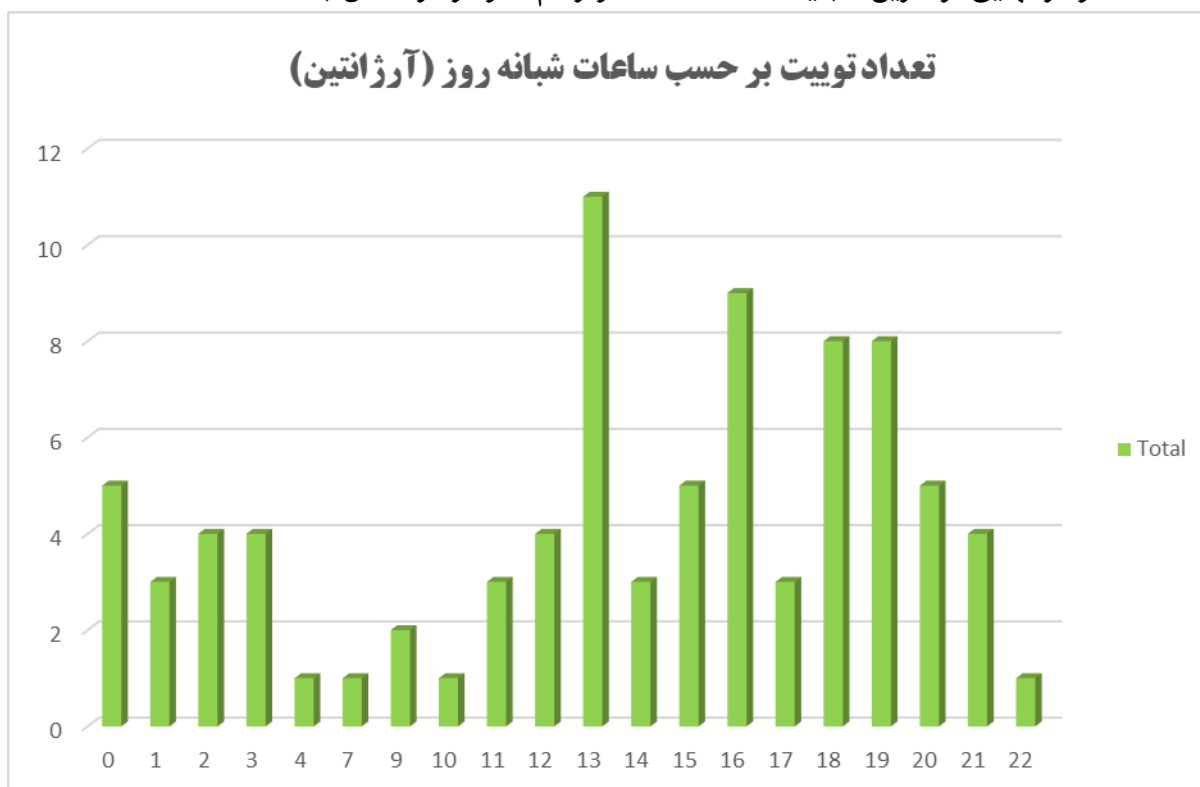
<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	---

۳/۴ گزارش شماره ۴

عنوان گزارش: ساعاتی از روز که بیشترین فعالیت در توییت وجود دارد به تفکیک کشور

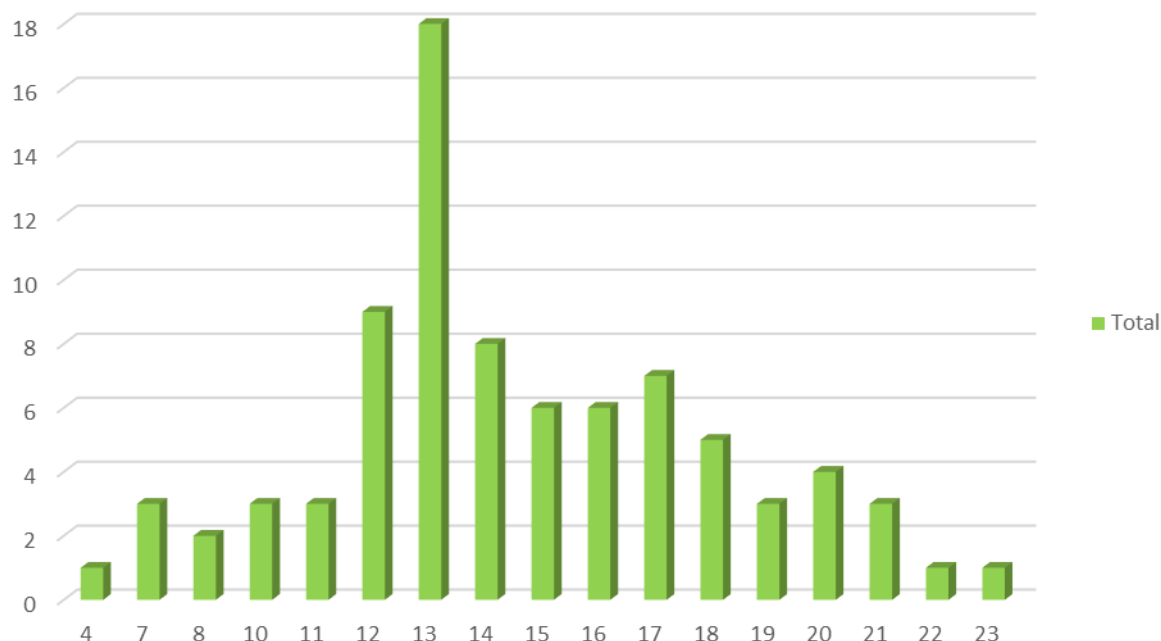
حجم این گزارش بسیار زیاد بوده و شامل حدوداً ۱۶۶ گزارش (به تعداد کشورهای موجود در توییت‌ها) می‌باشد. در این بخش تنها گزارش‌های ۳ کشور به عنوان نمونه ارائه گردیده است. (تمام گزارش‌ها در اکسل مربوطه در دسترس می‌باشد).

- نمودار نهایی از طریق قابلیت Pivot Table و رسم نمودار در اکسل بدست آمده است.

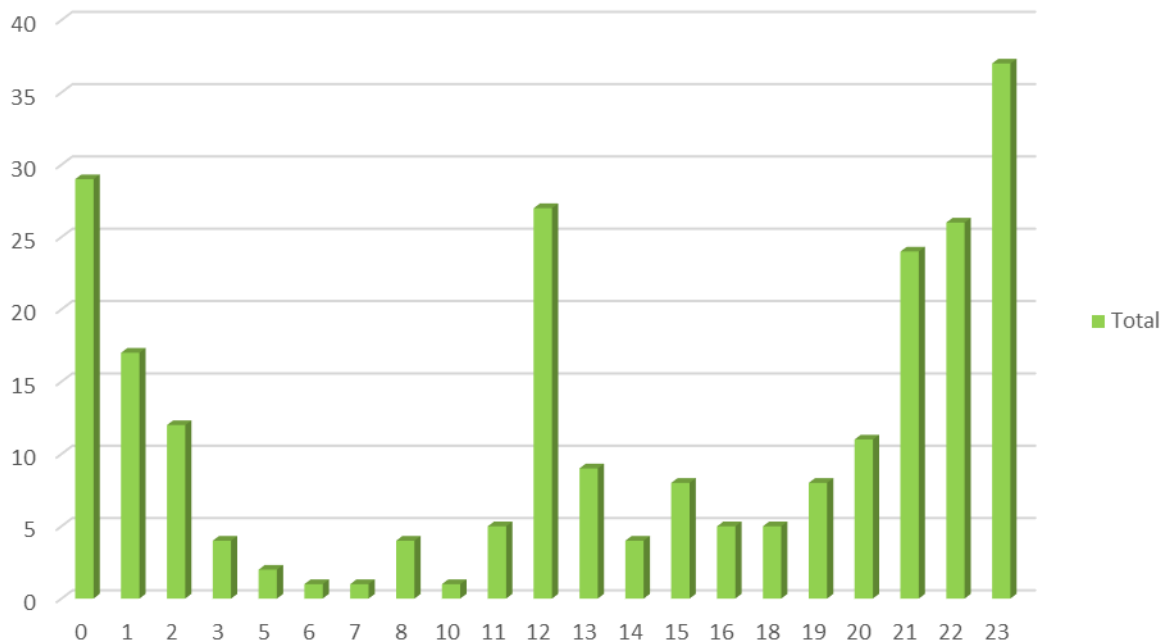




تعداد تویت بر حسب ساعات شبانه روز (روسیه)



تعداد تویت بر حسب ساعات شبانه روز (ژاپن)



<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

۳/۵ گزارش شماره ۵

عنوان گزارش: میزان مثبت یا منفی بودن (بار احساسی) توییت‌های منتشرشده

خروجی نهایی این گزارش تنها یک عدد بوده که نشان‌دهنده حس کلی موجود در توییت‌های مربوط به شرکت اپل می‌باشد. این گزارش می‌بایست به صورت آنی و به کمک الگوریتم‌های سالمندی تهیه گردد.

برای تهیه این گزارش مراحل زیر انجام پذیرفته است:

- حذف کاراکتر «خط جدید» از متن توییت‌ها
- استخراج متن توییت‌ها از اکسل و ذخیره آن در فایل متنی ساده (txt) {هر توییت در یک سطر}
- نوشتن کد پایتون برای تحلیل حساسیت متن توییت‌ها به کمک nltk
- اجرای کد به منظور دریافت خروجی نمرات مربوط به آنالیز حساسیت در قالب فایل متنی ساده (txt) {نمرات هر توییت در یک سطر}
- وارد کردن اطلاعات نمرات در فایل اکسل اولیه
- محاسبه میانگین نمره^۲ تمام توییت‌ها در اکسل

میزان مثبت بودن توییت‌های کاربران

7,849,176

^۱ منظور کاراکتر n می‌باشد.

^۲ نمره نهایی که در گزارش ارائه شده است، ۱۰۰ میلیون برابر نمره‌ی compound می‌باشد.

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

۳/۶ گزارش شماره ۶

عنوان گزارش: میزان مثبت یا منفی بودن (بار احساسی) توییت‌های منتشرشده به تفکیک کشور
حجم این گزارش بسیار زیاد بوده و شامل حدوداً ۱۶۶ گزارش (به تعداد کشورهای موجود در توییت‌ها) می‌باشد. در این بخش تنها گزارش‌های ۱۰ کشور به عنوان نمونه ارائه گردیده است. (تمام گزارش‌ها در اکسل مربوطه در دسترس می‌باشد).

میزان مثبت بودن توییت‌های کاربران (آرژانتین)
-327,412
میزان مثبت بودن توییت‌های کاربران (روسیه)
1,322,289
میزان مثبت بودن توییت‌های کاربران (ژاپن)
436,042
میزان مثبت بودن توییت‌های کاربران (آمریکا)
9,513,335
میزان مثبت بودن توییت‌های کاربران (فرانسه)
4,346,866
میزان مثبت بودن توییت‌های کاربران (گرجستان)
13,549,756
میزان مثبت بودن توییت‌های کاربران (اوکراین)
-4,308,824
میزان مثبت بودن توییت‌های کاربران (ایران)
9,866,667
میزان مثبت بودن توییت‌های کاربران (هند)
10,411,207
میزان مثبت بودن توییت‌های کاربران (کانادا)
6,574,709

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

۳/۷ گزارش شماره ۷

عنوان گزارش: میزان مثبت یا منفی بودن توییت‌های منتشرشده به تفکیک محصول

میزان مثبت بودن توییت‌های کاربران مربوط به محصول iphone
11,739
میزان مثبت بودن توییت‌های کاربران مربوط به محصول ipad
8,009,927
میزان مثبت بودن توییت‌های کاربران مربوط به محصول ipod
470,821
میزان مثبت بودن توییت‌های کاربران مربوط به محصول mac
3,261,653

- تعیین ارتباط توییت با محصول از طریق ذکر شدن نام محصول در متن توییت صورت گرفته است.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

۳/۸ گزارش شماره ۸

عنوان گزارش: متن ۲۰ توییت برتر

رتبه	متن توییت
۱	<p>Apple has \$233 billion in cash. It could buy all</p> <ul style="list-style-type: none"> —@NFL teams —@NBA teams —@MLB teams —@NHL teams <p>...and still have \$80 billion left. \$AAPL</p>
۲	<p>Read \$MRNJ #NEWS, \$HEMP & \$GRCU r above .01, that's where we r goin🔪 \$SPY \$MSFT \$SBUX \$SFOR \$VRX \$AAPL \$TSLA \$GOOG \$FB \$EURUSD \$USDJPY \$MLCG</p>
۳	<p>\$RXSF is leadin their sector accordin 2 @EdisonMediaCen \$AAPL \$SPY \$TSLA \$FB \$EURUSD \$ALK \$IBB \$EW \$AMZN \$GBPUSD \$GM https://t.co/LY2mHn755</p>
۴	<p>Philstockworld Top Trade Review \$AAPL \$MSFT #Dividends \$USO \$HOV \$TWTR -- https://t.co/JArXslm7CI https://t.co/kRR9ezhm9E</p>
۵	<p>Philstockworld Top Trade Review: \$AAPL \$ABX \$BA \$CAKE \$CMG \$DIS \$IBM \$GILD \$LL \$UNG \$SPY -- https://t.co/EX5SYjdwBC https://t.co/7FBZwVZ63v</p>
۶	<p>Monday's Oil Mess: Rent-A-Rebel Jacks up Prices into the Holiday \$USO \$AAPL #Earnings -- https://t.co/cGHB3WDKA8 https://t.co/JFZIBcom1n</p>
۷	<p>Meaningless Monday Market Movement! \$AAPL \$SQQQ #oil #Brexit https://t.co/j4lqg7E1HN</p>
۸	<p>S&P Futures Back over 2,050, for Now \$SPY \$AAPL \$SQQQ #China #Debt #Hedging -- https://t.co/2dOc5T89S3 https://t.co/TDPVdNRNQF</p>
۹	<p>Apple has \$233 billion in cash. It could buy:</p> <ul style="list-style-type: none"> Uber Tesla Twitter Airbnb Netflix Yahoo <p>...and still have \$18 billion left. \$AAPL</p>
۱۰	<p>🌟TURN YOUR \$500 INTO \$5,000+🌟</p> <p>JOIN #TEAMBILLIONAIRE👉</p> <p>✉ pennystockhotline@gmail.com</p> <p>#PENNYSTOCKS \$AAPL \$UVXY \$JDST https://t.co/lwAGjfmIP3</p>

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	---

۱۱	<p>✳️TURN YOUR \$500 INTO \$5,000+✳️</p> <p>JOIN #TEAMBILLIONAIRE ↗</p> <p>✉️ pennystockhotline@gmail.com</p> <p>#PENNYSTOCKS \$AAPL \$GSAT \$MGT</p> <p>https://t.co/lwAGjfmIP3</p>
۱۲	<p>Option Opportunity Portfolio May Review – Up 19.3% In 30 Days! \$ABX \$FCX \$USO \$AAPL \$DIS - https://t.co/rp3kMsRZ3E https://t.co/TKkc15pKcR</p>
۱۳	<p>Trendless Tuesday - Watch Yesterday's Fake Gains Disappear \$AAPL #China \$FXI #Earnings - https://t.co/GpgGqoOIFn https://t.co/FRuixv5aZF</p>
۱۴	<p>Tempting Tuesday - S&P 2,100 is Still the Line to Watch Ahead of the Fed \$AAPL \$QQQ -- https://t.co/t1eDfKHJnk https://t.co/BAW3RAe7SC</p>
۱۵	<p>Waiting for the Fed – Apple Gives Us Huge Wins: \$AAPL \$SQQQ #GDP #Nikkei #Futures #Oil -- https://t.co/Al3pkf350V https://t.co/LktIRF4F2b</p>
۱۶	<p>Our \$SQQQ Hedge is Up 314% and Our Futures Are Up \$4,850, You're Welcome! \$AAPL -- https://t.co/eUQ2kCkCOY https://t.co/Yk98oyqMZl</p>
۱۷	<p>TURN YOUR \$ 500 INTO \$ 5,000\$✳️</p> <p>JOIN #TEAMBILLIONAIRE ↗</p> <p>✉️ pennystockhotline@gmail.com</p> <p>#PENNYSTOCKS \$TWTR \$AAPL \$LNKD</p> <p>https://t.co/euJFNQX1g4</p>
۱۸	<p>TURN YOUR \$ 500 INTO \$ 5,000\$✳️</p> <p>JOIN #TEAMBILLIONAIRE ↗</p> <p>✉️ pennystockhotline@gmail.com</p> <p>#PENNYSTOCKS \$TALK \$PPPI \$AAPL https://t.co/oSn11kxftM</p>
۱۹	<p>Bears today. We getting paid! \$AAPL \$TWTR \$BWLD \$NFLX https://t.co/CCi0S3skJJ</p>
۲۰	<p>Apple has \$233 billion in cash. It could buy all</p> <p>—@NFL teams</p> <p>—@NBA teams</p> <p>—@MLB teams</p> <p>—@NHL teams</p> <p>...and still have \$80 billion left. \$AAPL</p>

- نحوه محاسبه امتیاز توییت‌ها به صورت زیر بوده است (فرمول ابتکاری):

$$\text{Score} = \text{Fav} \times \text{Retweet}^2$$

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	---

۳/۹ گزارش شماره ۹

عنوان گزارش: ۲۰ مورد از پرتکرارترین هشتگ‌ها

هشتگ	تعداد تکرار
#Apple	26516
#stocks	6277
#PENNYSTOCKS	3808
#TEAMBILLIONAIRE	3567
#investing	2502
#trading	2437
#Stocks	2270
#China	2101
#Futures	1899
#Earnings	1895
#stockmarket	1868
#Nikkei	1863
#NEWS	1829
#finance	1558
#APPLE	1510
#BREAKINGNEWS	1500
#WEEDSTOCKS	1307
#Dividends	1303
#MARIJUANASTOCKS	1300
#WEED	1298

- به منظور تهیه این گزارش ابتدا تمام داده‌ها به فایل متنی ساده (txt) تبدیل شده و سپس با اجرای یک قطعه کد پایتون تحلیل مورد نیاز صورت گرفته و خروجی به فایل اکسل اصلی ضمیمه شده است.

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	---

۳/۱۰ گزارش شماره ۱۰

عنوان گزارش: ۲۰ مورد از پر تکرارترین هشتگ‌ها به تفکیک کشور

حجم این گزارش بسیار زیاد بوده و شامل حدوداً ۱۶۶ گزارش (به تعداد کشورهای موجود در توییت‌ها) می‌باشد. در این بخش تنها گزارش‌های ۳ کشور به عنوان نمونه ارائه گردیده است. (تمام گزارش‌ها در اکسل مربوطه در دسترس می‌باشد).

پرتکرارترین هشتگ‌ها در ایالات متحده آمریکا	
هشتگ	تعداد تکرار
#NEWS	1386
#Futures	1047
#Nikkei	1046
#China	1020
#Earnings	1011
#Apple	892
#PENNYSTOCKS	799
#TEAMBILLIONAIRE	790
#Dividends	766
#Stocks	697
#stocks	577
#Productivity	549
#Debt	545
#Hedging	545
#Quant	537
#Oil	500
#GDP	498
#Invest	396
#YouAreWelcome	327
#trading	226
پرتکرارترین هشتگ‌ها در کشور فرانسه	
هشتگ	تعداد تکرار
#Apple	32
#NEWS	26
#TEAMBILLIONAIRE	26
#PENNYSTOCKS	26
#STOCKMARKET	10
#EmergingMarket	7
#stocks	5
#STOCKS	5
#Nikkei	4

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر</p>
---	--	--

#Futures	4
#iPhone	4
#TRADING	3
#Stocks	3
#tech	3
#Productivity	3
#CarPlay	3
#Bourse	3
#Earnings	2
#Oil	2
#finance	2
پرتکرارترین هشتگ‌ها در کشور آلمان	
هشتگ	تعداد تکرار
#Apple	99
#BUY	32
#aktien	19
#TEAMBILLIONAIRE	18
#PENNYSTOCKS	18
#iPhone	18
#DowJones	17
#HOLD	14
#STOCKMARKET	9
#AAPL	8
#SELL	8
#stocks	6
#iPhone7	5
#AppleCar	5
#GD200	5
#STOCKS	4
#AppleWatch	4
#Handelsblatt	4
#Nikkei	3
#trading	3

استاد درس: دکتر جعفر حبیبی نیمسال اول ۱۳۹۶-۱۳۹۷	پروژه درس معماری نرم افزار آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	 دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر
--	---	--

۳/۱۱ گزارش شماره ۱۱

عنوان گزارش: لیستی ۱۰ زبان پر استفاده برای توییت

تعداد توییت	زبان	کد ایزو
159864	انگلیسی	en
4871	اسپانیایی	es
683	هلندی	nl
456	آلمانی	de
249	فرانسوی	fr
165	فنلاندی	fi
103	پرتغالی	pt
79	دانمارکی	da
59	سوئدی	sv
59	روسی	ru

- جدول نهایی از طریق قابلیت Pivot Table در اکسل بدست آمده است.

۳/۱۲ گزارش شماره ۱۲

عنوان گزارش: لیستی ۱۰ کشوری که بیشترین فعالیت را در توییت دارند

تعداد توییت	کشور	کد ایزو
31186	ایالات متحده آمریکا	US
16080	بریتانیا	GB
2183	کانادا	CA
845	هند	IN
633	آلمان	DE
574	ایرلند	IE
562	جرسی (کشوری جزیره‌ای نزدیک انگلستان)	JE
552	فرانسه	FR
478	اسپانیا	ES
434	استرالیا	AU

- جدول نهایی از طریق قابلیت Pivot Table در اکسل بدست آمده است.

استاد درس: دکتر جعفر حبیبی	پروژه درس معماری نرم افزار	
نیمسال اول ۱۳۹۶-۱۳۹۷	آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت	دانشگاه صنعتی شریف دانشکده مهندسی کامپیوتر

ضمایم

ضمیمه شماره ۱: هرم DIKW و تفاوت سطوح مختلف [9]

داده: داده خام است و به خودی خود به جز موجودیتش خاصیتی ندارد، می‌تواند به هر شکلی (عدد، حروف، علائم و ...) وجود داشته باشد. قابل استفاده یا غیر قابل استفاده، و معنای خاصی ندارد. مثلاً a یک داده است.

اطلاعات: همان داده است که با ایجاد ارتباطات به آن معنا داده شده است، این «معنا» می‌تواند مفید باشد اما لزومی ندارد که اینگونه باشد. مثلاً $b > a$ (اینکه b بزرگتر از a است می‌تواند مفید نباشد)

دانش: مجموعه‌ای مناسب از اطلاعات است و به این شکل مفید نیز واقع می‌شود. دانش یک فرآیند کاملاً قطعی است. وقتی شخصی اطلاعات را به خاطر می‌سپارد در حال گردآوری دانش است این دانش برای آنها مفید است اما هنوز به یکپارچگی و قوام لازم دست نیافته است که بتواند برای آینده مورد استفاده قرار بگیرد به عنوان مثال فرض کنید دانش آموز کلاس چهارم ضرب را یاد گرفته بنابراین وقتی از او پرسید 4×4 او پاسخ می‌دهد اما اگر از او پرسید 44×44 نمی‌تواند پاسخ دهد، زیرا این مطلب هنوز در حوزه او قرار نمی‌گیرد و باید به کلاس پنجم برود پس هنوز به قوام لازم برای هر ضربی دست نیافته است. در قالب مثال قبل $b < c$, $a < b$ مجموعه این روابط دانش است.

بینش/درک: درک کردن یک فرآیند منقطع و احتمالی است، یک فرآیند شهودی و تحلیلی، فرآیندی که به کمک آن می‌تواند دانش قبلی که در ذهن خود دارد را به دانش جدید تبدیل کند. تفاوت بین درک و دانش تفاوت بین به خاطر سپردن و یادگرفتن است افرادی که به درک کافی می‌رسند می‌توانند اقدامات مناسبی را اتخاذ کنند زیرا می‌توانند دانش جدیدی را بسازند یا در برخی موارد حداقل اطلاعات جدید، بر اساس مطالبی که قبلاً یادگرفته‌شده است، درک می‌گردد. برای نمونه:

$b > a$, $c > b$ می‌دهد که $c > a$ که یک دانش جدید است که از دانش قبلی به درک رسید.

خرد: خرد یک فرآیند غیر قطعی و غیر مجتمع و غیر احتمالی است. خرد در برابر سطوح قبلی آگاهی و به طور خاص در برابر گونه‌های ویژه‌ای از برنامه‌ریزی ماهیت انسانی (اخلاقی، رفتاری و ...) مطرح می‌شود. خرد همانند چراغ راهی است که به ما راه را نشان می‌دهد. بر اساس دانش‌های قبلی شکل گرفته اما فراتر از درک می‌باشد و دانشی را به ما ارائه می‌دهد که تا قبل از این نداشته‌ایم. اساس پرسشگری‌های فلسفی، خرد است. و بر خلاف ۴ سطح قبلی سوال‌هایی را می‌پرسد که برای آنها پاسخ ساده‌ای وجود ندارد و در برخی موارد سوالاتی که بشر از پاسخ به آنها عاجز است.

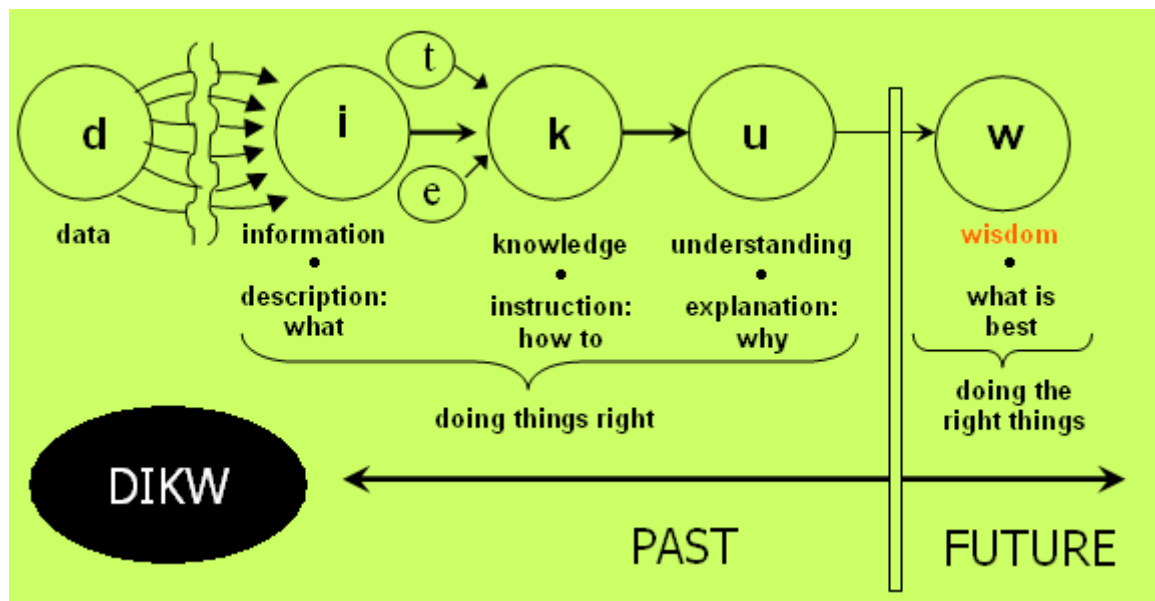
<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

این هرم DIKW نام دارد و سطوح و وسعت هر کدام از موارد بحث شده را نشان می‌دهد:



شکل ۴-۱ هرم DIKW

اگر بخواهیم جایگاه سطوح داده تا خرد را در محور زمان نشان دهیم و نتایج را ببینیم شکل زیر را خواهیم داشت:



شکل ۴-۲ جایگاه سطوح مختلف DIKW در محور زمان

راهنمای اختصارات شکل:

d=data, i=information, k=knowledge, u=understanding, w=wisdom, t=tacit knowledge, e=explicit knowledge

<p>استاد درس: دکتر جعفر حبیبی</p> <p>نیمسال اول ۱۳۹۶-۱۳۹۷</p>	<p>پروژه درس معماری نرم افزار</p> <p>آشنایی با معماری سامانه‌های نرم افزار به عنوان خدمت</p>	 <p>دانشگاه صنعتی شریف</p> <p>دانشکده مهندسی کامپیوتر</p>
---	--	--

مراجع

- [1] F. Dweiri, S. Kumar, S. Ahmed Khan and V. Jain, "Designing an integrated AHP based decision support system for supplier selection in automotive industry," vol. 62, pp. 273-283, 2016.
- [2] N. Vyas, K. Paithankar and S. Joshi, "A Novel Approach for Design of Ontology Based Clinical Decision Support System," in *International Journal of Advanced Research in Computer Science*, 2017.
- [3] "Data Warehouse," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Data_warehouse.
- [4] "Cassandra," Apache, [Online]. Available: <http://cassandra.apache.org/>.
- [5] "fasttext," Facebook, [Online]. Available: <https://fasttext.cc/>.
- [6] "Natural Language Toolkit," [Online]. Available: <https://www.nltk.org/>.
- [7] "Regression Analysis," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Regression_analysis.
- [8] "C4.5 Algorithm," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/C4.5_algorithm.
- [9] "DIKW Pyramid," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/DIKW_pyramid.