



دانشگاه صنعتی امیر کبیر  
(پلی تکنیک تهران)

"یا علام"

گزارش فاز اول

---

پژوهش، طراحی و پیاده‌سازی سامانه نرم‌افزاری داده‌کاوی و  
متن‌کاوی بر روی داده‌های مرکز پژوهش و سنجش افکار  
صداوسیما با قابلیت پیش‌بینی رفتار مخاطبان

فاز اول: نیازسنجی‌های پروژه

---

نسخه: ۱

تاریخ آخرین تغییرات: ۱۳۹۷/۱۱/۹

دانشکده مهندسی کامپیوتر و فناوری اطلاعات  
دانشگاه صنعتی امیر کبیر





## فهرست مطالب

چکیده	۶
شناسنامه گزارش	۷
۱. روشن کردن صورت مسئله	۸
۱,۱. ارزیابی زیرساخت کنونی	۸
۱,۲. ساختار پروژه	۸
۱,۳. گزارش حوزه چشم‌انداز	۹
۲. برنامه ریزی	۱۳
۲,۱. برنامه‌ها	۱۳
۲,۱,۱. طرح دسترس‌پذیری	۱۳
۲,۱,۲. طرح Recovery و Backup	۱۳
۲,۱,۵. طرح استقرار	۱۳
۲,۱,۶. طرح توسعه	۱۳
۲,۱,۷. طرح پشتیبانی کاربر نهایی	۱۵
۲,۱,۸. طرح جامع پروژه	۱۶
۲,۱,۹. طرح پشتیبانی	۱۸
۲,۱,۱۰. طرح مانیتورینگ	۱۸
۲,۱,۱۱. طرح عملیات	۱۹
۲,۱,۱۲. طرح کارایی	۱۹
۲,۱,۱۳. طرح پایلوت	۱۹
۲,۱,۱۴. طرح امنیت	۲۰
۲,۱,۱۶. طرح آزمون	۲۰
۲,۲. مشخصات	۲۱
۲,۲,۱. نیازمندی‌های کسب و کار	۲۱
۲,۲,۲. مشخصات عملکردی	۲۱
۲,۲,۳. طرح منطقی	۲۶

۲۶	..... طرح مفهومی	۲,۲,۴
۲۹	..... نیازمندی‌های عملیاتی	۲,۲,۵
۲۹	..... طرح فیزیکی	۲,۲,۶
۲۹	..... سناریوهای استفاده	۲,۲,۸
۲۹	..... نیازمندی‌های کاربر	۲,۲,۹

## چکیده

مستند حاضر گزارش فاز اول پروژه «پژوهش، طراحی و پیاده‌سازی سامانه نرم افزاری داده‌کاوی و متن‌کاوی بر روی داده‌های مرکز پژوهش و سنجش افکار صداوسیما با قابلیت پیش‌بینی رفتار مخاطبان» می‌باشد که در آن به بررسی نیازسنجی سیستم پرداخته شده‌است.

هدف اصلی این پروژه ارائه سیستمی برای تحلیل خودکار نظرات افرادی است که در نظر سنجی‌های صدا و سیما شرکت نموده‌اند و به سؤالات چند گزینه‌ای و همین‌طور سوال پیشنهاد و نظر آزاد پاسخ داده‌اند. در این تحلیل ارتباط بین پاسخ‌های افراد به سؤالات و پرسشنامه‌های مختلف مورد بررسی قرار خواهد گرفت. این تحلیل می‌تواند به صورت کلی برای تمام پرسش‌شوندگان انجام گیرد و یا بر روی گروه‌های مشخصی که افراد به آنها تعلق دارند، انجام شود. به عنوان مثال افراد متعلق به یک منطقه جغرافیایی خاص یا افراد دارای مدرک تحصیلی مشخص.

انجام این پروژه و ارائه یک تحلیل جامع از نظرات افراد جامعه می‌تواند نقش به‌سزایی در یافتن دغدغه‌های اصلی مردم و میزان و دلیل رضایتمندی و یا نارضایتی آنها از موارد مورد بررسی و توجه صدا و سیما داشته باشد که به تبع آن ارائه راهکارهای مفید در رفع کاستی‌ها میسرتر خواهد بود.

شایان ذکر است با توجه به اینکه پروژه حاضر دارای یک پروژه تکمیلی آتی برای گسترش سیستم می‌باشد. با توجه به اینکه زیرساخت‌های پروژه آتی باید در این سیستم مد نظر قرار گیرد تا بتوان به قابلیت‌های مد نظر برای سیستم گسترش یافته رسید، در مستند حاضر به آن قابلیت‌ها نیز اشاره شده است و زیرساخت مورد نیاز در زمان اجرا در نظر گرفته خواهد شد.

## شناسنامه گزارش

شناسنامه گزارش		
پژوهش، طراحی و پیاده‌سازی سامانه نرم افزاری داده‌کاوی و متن‌کاوی بر روی داده‌های مرکز پژوهش و سنجش افکار صداوسیما با قابلیت پیش‌بینی رفتار مخاطبان		نام پروژه
۱۹۰۰/۲۲۸۶۱		شناسه قرارداد
دانشگاه صنعتی امیرکبیر		مجری پروژه
۱		فاز پروژه
گزارش نیازسنجی		نام فاز
اول	شماره نسخه	نسخ تهیه‌شده برای این فاز
۲	تعداد فصل	
۲۶	تعداد صفحات	اطلاعات گزارش

## ۱. روشن کردن صورت مسئله

**۱.۱. ارزیابی زیرساخت کنونی :** سیستم کنونی بدین صورت است که تحلیلگرها به صورت دستی با استفاده از نرم افزار SPSS، تست‌های آماری را گرفته و در نهایت تحلیل‌های نهایی سه بعدی خود که شامل تحلیل‌های شمارشی درصد افراد گروه مختلف است را از فایل‌های SAV که توسط اپراتورهای سازمان از روی فرم‌های نظرسنجی پر شده است، برداشت کرده به صورت گزارش کتبی ارائه می‌کنند. در این سیستم به هیچ وجه به ارتباط بین نظرسنجی‌ها توجه نمی‌شود. با توجه به اینکه داده‌های کتبی نیز به صورت کدگذاری شده به دست تحلیلگرها می‌رسد امکان تحلیل متون نیز وجود ندارد و با آن‌ها همچون داده‌های چندگزینه‌ای برخورد می‌شود. چون این تحلیل درون سازمان انجام می‌شود و به صورت کتبی نتیجه‌اش اعلام می‌شود و دارای سیستمی نیست که نتایج به صورت یکجا و آنلاین و یا از بیرون از سازمان دیده شود و به همین دلیل نیاز به امنیت خاصی هم ندارد. در حالی که در سیستم جدید داده‌های به صورت یکپارچه و مجتمع در پایگاه‌های داده نگهداری می‌شود و با استفاده از وب اپلیکیشن به داده‌ها و تمام تحلیل‌های متون و داده‌ها و ارتباط بین پرسشنامه‌ها می‌توان دسترسی پیدا کرد و کارشناسان ارشد و مدیران می‌توانند نتایج تحلیل را در داشبوردی به راحتی مشاهده و بررسی کنند. این سامانه برای تحلیلگران در داخل سازمان بکار گرفته می‌شود و مدیران نیز هم در شبکه داخلی و هم از طریق اینترنت به نتایج تحلیل دسترسی خواهند داشت. امنیت سامانه به واسطه استفاده از چارچوب‌های برنامه نویسی استاندارد، رعایت کردن چک لیست‌های امنیتی و در نهایت تست‌های نفوذ توسط افراد خبره حوزه امنیت به خوبی رعایت می‌شود.

**۱.۲. ساختار پروژه :** این پروژه ارائه سیستمی با هدف ارائه گزارش‌ها و متغیرهای آماری، تحلیل داده‌های چهارگزینه‌ای و تحلیل متن‌های نظرسنجی است. این تحلیل ارتباط بین پاسخهای افراد به سوالات مختلف مورد بررسی قرار می‌دهد و میتواند به صورت کلی برای تمام پرسششوندگان و یا بر روی گروههای مشخصی که افراد به آنها تعلق دارند، انجام شود. این تحلیل‌ها به خوبی می‌تواند کمک کند که بتوان مواردی همچون شناختن هرچه بیشتر سلايق، دغدغه‌ها، نظرات و میزان رضایتمندی مردم از موارد مورد بررسی صدا و سیما، سنجش میزان اثربخشی محتوای تولیدشده، شناسایی رفتار مخاطبان و آشنایی با نیازها و خواسته‌های آن‌ها، پیش‌بینی رفتار مخاطبان مثلاً در مناسبت‌های سیاسی مانند انتخابات، تجمیع و تحلیل یکپارچه نظرسنجی‌های صورت گرفته از حوزه‌های مختلف بهتر اتفاق بیفتد.

در این سامانه در فاز اول فایل‌هایی با فرمت‌های sav، csv و یا جداول پایگاه داده‌های SQL server<sup>۱</sup> دریافت و اطلاعات مورد نیاز در پایگاه داده‌های سامانه ذخیره می‌شود. سپس پیش پردازش روی داده‌ها انجام می‌گیرد که در متون شامل جداسازی کلمات، یکسان‌سازی، نرمال‌سازی و حذف ایست‌واژه‌ها و در داده‌ها شامل حذف نویز، حذف داده‌های پرت و تصمیم‌گیری درباره‌ی سوالاتی بی‌پاسخ است. این سامانه گزارشات آماری را ارائه می‌دهد که شامل ارائه انواع گزارشهای آماری از پاسخ نامیها، بررسی همبستگی میان پاسخهای سوالات مختلف در یک پرسشنامه و ارائه انواع

<sup>۱</sup> بخش پایگاه داده SQL server در پروژه گسترش یافته اضافه خواهد شد



آزمونهای آماری است. داده‌هایی که این سیستم پس از تحلیل داده‌ها ارائه می‌دهد شامل ارائه الگوهای تکرار شونده پنهان در پاسخنامه‌ها، پیش‌بینی رفتار افراد بر اساس الگوهای استخراج شده، یافتن افراد مشابه هم در جامعه و بیان مهمترین ویژگیهای آنها و دسته‌بندی داده‌ها بر اساس ویژگیهای مشابه است. متن کاوی توسط این سامانه نیز به استخراج موضوعات اصلی نظرات متنی، یافتن علایق و نظرات گروه‌های مختلف شرکت‌کننده در نظر سنجی بر مبنای تفکیک رده سنی، جنسیت، میزان تحصیلات و محل سکونت، تحلیل افکار مطرح حول هر موضوع در پاسخهای پرسشهای آزاد از لحاظ مثبت یا منفی بودن منجر می‌شود. در این پروژه مواردی همچون چگونگی ایجاد داده‌های SAV و CSV توسط اپراتورها، طرح آموزش، تحلیل‌ها و متغیرهایی آماری غیر از موارد مذکور در ادامه دیده نشده است. همانطور که مشخص است این سیستم نسبت به روش قبلی مزیت‌های زیادی دارد که می‌توان به تحلیل‌های کامل پرسشنامه‌ها، تحلیل اطلاعات روابط بین پرسشنامه‌ها<sup>۱</sup>، تحلیل دقیق متون نظرسنجی به جای تبدیل آن‌ها به کدها توسط اپراتورها و برخورد با آنها مانند نظرسنجی‌های چند گزینه‌ای، افزایش سرعت در تحلیل‌ها، تحلیل داده‌ها در سطوح مختلف، یکجاسازی و در دسترس بودن تحلیل‌ها با وب اپلیکیشن در کل شبکه داخلی و مشاهده نتایج تحلیل از بیرون از سازمان بدون نگرانی درباره امنیت داده‌ها اشاره کرد.

**۱.۳. گزارش حوزه چشم‌انداز:** در این پروژه گزارشات آماری و تحلیل جامعی از داده‌های متنی و چندگزینه‌ای و ارتباط بین پرسشنامه‌ها<sup>۲</sup> ارائه می‌شود که منجر به یکپارچه‌سازی و مجتمع‌سازی داده‌ها و تحلیل‌ها در یک پایگاه داده و سیستم می‌شود که در آینده نیز می‌توان به یکپارچه‌سازی کل سیستم از ابتدای وارد کردن داده‌ها توسط اپراتورها و در ادامه یکپارچه‌سازی به وارد کردن مستقیم داده‌ها توسط پرسش‌کنندگان میدانی، اتصال به پایگاه‌های اینترنتی که نظرسنجی در آن‌ها انجام می‌شود و سیستم پیامکی سازمان روی آورد. برای طراحی این سامانه نیاز به نمونه‌های مختلفی است که عبارت است از داده‌های با فرمت‌های SAV، CSV و schema در جداول SQL<sup>۳</sup> و چندین سطر از جداول آن برای آشنایی با ساختار داده‌ها، چند نمونه از تحلیل‌هایی که تحلیلگران به صورت دستی انجام می‌دادند، متن‌های نظرسنجی به اندازه کافی برای تحلیل‌های متن و نمونه کد شده آن‌ها توسط اپراتورها. همچنین برای آشنایی با سلیقه و عادات کاربری و شبکه داخلی سازمان و سطوح دسترسی، نیاز است با کارشناسان تحلیلگر و مدیران استفاده کننده از تحلیل صحبت شود و در ادامه نیز صحبت با آن‌ها در فازهای مختلف و دریافت پیشنهادهای تجارب کاربری آن‌ها در استفاده از سیستم جدید لازم، ضروری و بهبود دهنده کیفیت خواهد بود. در این سیستم نیازها و عملکردهای زیر برآورده خواهد شد:

• مدیریت پرسشنامه‌ها

- مشاهده لیست پرسشنامه‌ها، با امکان جستجو و فیلتر بر اساس ویژگی‌هایی مانند عنوان، تاریخ ایجاد، تاریخ قفل و متادیتا
- مشاهده تعداد کل پاسخ‌ها و آمار کلی در مورد هر پاسخ‌نامه

<sup>۱</sup> بخش تحلیل روابط بین پرسشنامه‌ها در پروژه گسترش یافته اضافه خواهد شد  
<sup>۲</sup> بخش تحلیل روابط بین پرسشنامه‌ها در پروژه گسترش یافته اضافه خواهد شد  
<sup>۳</sup> بخش پایگاه داده SQL server در پروژه گسترش یافته اضافه خواهد شد

- قابلیت مشاهده و ویرایش اطلاعات اضافی دیگر (متادیتا) مانند جامعه آماری (به صورت متنی)، تاریخ اجرا، شیوه اجرا (تلفنی، میدانی، اینترنتی) نیز موجود باشد که در ادامه جهت دسته‌بندی و گزارش‌گیری از این داده‌های افزوده استفاده شود.
- دسته‌بندی موضوعی و ویرایش پرسشنامه‌ها
- حذف، ویرایش و قفل کردن پرسشنامه‌ها
- ارتباط دهی نظرسنجی‌های مختلف و یکسان‌سازی پارامترهای مشابه ولی بانام‌های مختلف در نظرسنجی‌ها

• گزارشات آماری

- امکان انتخاب و جستجوی یک یا چند پرسشنامه‌ها جهت انجام گزارشات آماری
- جدول‌های فراوانی تک‌تک پرسش‌ها شامل فراوانی، درصد و درصد تراکمی
- جدول‌های متقاطع بین پرسش‌ها (حداکثر ۵ مورد) با امکان انتخاب فراوانی خانه‌ها، درصد کل، درصد سطری، درصد ستونی و درصدهای حاشیه‌ای
- محاسبه مشخص‌های توصیفی (میانگین، میانگین پیراسته، انحراف معیار، دامنه، مقدار کمینه، مقدار بیشینه، میانه، نما، چندک‌ها، چولگی و کشیدگی)
- محاسبه نمره کل از مجموعه‌ای از پرسش‌ها و تعیین دامنه نمره کل با نمره کل به صورت وزنی یا غیر وزنی و انتخاب بازه و شیوه محاسبه شاخص
- محاسبه ضرایب همبستگی پیرسون، اسپیرمن، تاو کندال، وی کرامر، دی سامرز و ضریب توافقی و آزمون‌های آماری آن‌ها
- محاسبه شاخص‌های اِتا دو و اِتا دو
- نمایش جدول و رسم نمودارهای دایره‌ای، میله‌ای، هیستوگرام، پراکنش و خطی با امکان چاپ
- اجرای آزمون‌های آماری خی‌دو، تی، اف (تجزیه واریانس)، کروسکال‌والیس، من‌ویتنی، میانه
- اجرای آزمون گردش (تصادفی بودن نمونه، Run Test)
- محاسبه واریانس درون بلوکی و بین بلوکی و همبستگی درون خوشه‌ای (بلوکی)
- اجرای تمام محاسبات ذکرشده در زیرگروه‌ها
- ارائه خروجی‌های فارسی (به‌ویژه ارقام، درصدها و نتایج آزمون‌ها) با قابلیت استفاده در مجموعه Excel و Word
- تعریف پرسش تفکیک‌کننده جهت تفکیک گزارش‌های آماری
- ارائه تمام روش‌های سنجش همبستگی از قبیل مدل معادلات ساختاری، تحلیل عاملی (تحلیلی و اکتشافی) و آزمون‌های مرتبط شامل آزمون کرویت و KMO، محاسبه ضریب پایایی با انتخاب مجموعه‌ای از سؤالات به انتخاب کاربر، محاسبه ضریب همبستگی چند رشته‌ای و چند رشته‌ای نقطه‌ای با انتخاب مجموعه‌ای از سؤالات به انتخاب کاربر
- آزمون نرمالیتی (انتخاب یک سؤال و بررسی نرمال بودن مقادیر توزیع‌شده)

- خوشه‌بندی بر اساس ویژگی‌های انتخابی
- export پاسخنامه‌هایی که باهم تجمیع شده در قالب سؤالات ادغام‌شده به‌صورتی که پرسشنامه هر رکورد در فایل تجمیعی مشخص شده باشد.
- جان‌هی (imputation) افراد دارای پرسش بی‌پاسخ به روش‌های نزدیک‌ترین همسایه، جانشینی پرسش همبسته، جانشینی با انتخاب تصادفی، جان‌هی رگرسیونی، جان‌هی نسبتی و جان‌هی میانگین به سه شکل میانگین فرد، میانگین پرسش و میانگین پرسش تصحیح‌شده

#### • داده‌کاوی

- بررسی همبستگی میان پاسخ‌های سؤالات مختلف در یک پرسشنامه و یا در چند پرسشنامه مرتبط به هم، محاسبه همبستگی (اعم از اسپیرمن و پیرسان) بین پاسخ‌های سؤالات. این ارتباط بین نظرات در پرسشنامه‌ها در صورتی می‌تواند تحلیل گردند که برای موارد مشابه در زمان‌های مختلف از نام یکسانی استفاده شده باشد.
- ارائه الگوهای تکرارشونده (ترکیبات چندتایی از پاسخ‌ها که درصد بالایی از پرسشنامه‌ها وجود دارد) پنهان در نتایج نظرسنجی‌ها
- یافتن افراد مشابه هم در جامعه و بیان مهم‌ترین ویژگی‌های آن‌ها با گروه‌بندی آنها بر اساس نحوه پاسخ به پرسش‌های مختلف با استفاده از الگوریتم‌های خوشه‌بندی، بررسی ویژگی‌های آماری این گروه‌ها و انتخاب حداقل سه الگوریتم خوشه‌بندی مناسب
- پیش‌بینی رفتار افراد بر اساس الگوهای استخراج‌شده و پاسخ به پرسشی که فرد به آن جواب نداده است پس از خوشه‌بندی و گروه‌بندی براساس پاسخ‌های مختلف
- پیش‌بینی قوانین و روابط مختلف با دقت حداقل ۹۰٪، محاسبه ضریب اطمینان و نرخ مثبت کاذب کمتر از ۰٫۱ برای هر همبستگی هریک از روابط، امکان تنظیم دقت گزارشات توسط کاربر
- قابلیت ارائه خروجی‌های فارسی گزارشات داده‌کاوی سامانه باقابلیت استفاده در مجموعه Excel و Word.
- امکان تولید قالب گزارشی جهت ارائه در داشبورد مدیریتی

#### • متن‌کاوی

- استخراج موضوعات اصلی نظرات متنی : استخراج موضوعات کلیدی کل متن و نمایش ۲ تا ۱۰ کلمه کلیدی از هر موضوع برای تبیین آن
- تحلیل افکار مطرح حول هر موضوع در پاسخ‌های پرسش‌های آزاد از لحاظ مثبت یا منفی بودن : پس از استخراج موضوعات کلیدی سیستم با تحلیل نظرات مثبت و منفی حول هر موضوع علاوه بر کلمات کلیدی که مبین موضوع هستند یک امتیاز مثبت و یک امتیاز منفی نیز به آن موضوع اختصاص می‌دهد. این کار باعث می‌شود بجای اینکه از کل نظرات موجود در یک نظرسنجی، فقط مثبت یا منفی بودن نظرات کلی مشخص شود، به تفکیک موضوع، مثبت یا منفی بودن به همراه مقدار موارد مشخص شود.

- یافتن علایق و نظرات گروه‌های مختلف شرکت‌کننده در نظرسنجی بر مبنای تفکیک رده سنی، جنسیت، میزان تحصیلات و محل سکونت و کل ویژگی‌های موجود در نظرسنجی‌ها : کاربر می‌تواند در این بخش یکی از ویژگی‌های داده‌ها را مشخص کند و متون باز را بر اساس آن ویژگی فیلتر نماید. به‌طورکلی این امکان در سیستم وجود خواهد داشت که بتوان افراد را بر مبنای یک ویژگی خاص انتخاب نماید و برای آن زیرمجموعه از افراد بررسی آماری روی پاسخ‌ها (که بیانگر نظرات و علایق افراد است) صورت می‌گیرد.
- امکان تولید قالب گزارشی جهت ارائه در داشبورد مدیریتی

• داشبورد گزارشات (داشبورد مدیریتی)

- قابلیت نمایش تمام گزارشات ایجاد شده در بخش‌های قبل با توجه به نوع دسترسی
- قابلیت شخصی سازی داشبورد توسط کاربر نهایی
- دسترسی به اطلاعات اصلی پرشنامه‌ها
- ساده بودن و درک سریع بصری
- قابلیت طراحی انواع نمودارهای متنوع آماری Pie ، میله ای و..
- قابلیت نمایش گزارشات به صورت جدولی و نموداری
- قابلیت تغییر رنگ و فونت و اندازه گزارش با سلیقه کاربر
- قابلیت نمایش در موبایل

## ۲. برنامه ریزی

### ۲.۱. برنامه‌ها

**۲.۱.۱. طرح دسترس‌پذیری :** در این قسمت باید از نظر سخت افزاری و نرم افزاری به مبحث دسترسی پذیری پرداخته شود که شامل backup و recovery، تست، آموزش، استقرار و مانیتورینگ است که هر یک از این‌ها در بخش خود توضیح داده می‌شوند و بخش سخت‌افزاری نیز شامل امکانات سخت افزاری مانند روتر و اینترفیس‌های شبکه لازم در سازمان است که در این گزارش گنجانده نمی‌شود.

**۲.۱.۲. طرح Recovery و Backup :** در این سیستم با استفاده از schedule task‌هایی که روی سرور قرار داده می‌شود که در بازه‌های زمانی تعیین شده (مثلاً یک روز یکبار) از تمام پایگاه‌های داده نسخه‌های پشتیبان با تاریخ ذخیره می‌شود و همچنین در صورت نیاز از پکیج‌های چارچوب لاراول همچون backup و laravel-backup برای بازیابی آن‌ها استفاده می‌شود. با توجه به اینکه این سیستم دارای transaction‌های کوچک با پایگاه داده نیست بنابراین با ذخیره داده‌ها به صورت منظم یکپارچگی داده حفظ خواهد شد و نگرانی از این بابت نیز وجود ندارد.

**۲.۱.۵. طرح استقرار:** این سیستم نیازمند نصب زبان پایتون و کتابخانه‌های لازم تحلیل متن و داده، زبان php برای استفاده از چارچوب لاراول و وب سرور nginx می‌باشد که باید روی سرور نصب شود. همچنین نصب آخرین ورژن مرورگر روی کامپیوترهای استفاده کنندگان نیز می‌تواند به استفاده هر چه بهتر از سیستم منجر شود. با توجه به اینکه تا قبل از این سیستم دیگری استفاده نشده است برای فاز ابتدایی، گذار پیش نمی‌آید و برای فازهای بعدی نیز با توجه به ساختار پروژه و مازول بندی شدن آن در حین استفاده از آن مشکلی برای اضافه شدن بقیه اسکریپت‌های پایتون و php پیش نخواهد آمد.

**۲.۱.۶. طرح توسعه :** این سیستم شامل یک وب اپلیکیشن برای تعامل با کاربران و مازول‌های تحلیل داده و متن است. بخش وب اپلیکیشن شامل API، یک پایگاه داده و بخش UI است. برای گسترش سمت سرور وب اپلیکیشن از چارچوب لاراول و برای سمت کاربری نیز از html و css برای گسترش استفاده می‌شود. همچنین از bootstrap برای responsive کردن وب اپلیکیشن در دستگاه‌ها با ابعاد مختلف و از کتابخانه jQuery برای داینامیک کردن استفاده خواهد شد. مازول‌های محاسبات آماری، تحلیل متن و تحلیل داده نیز با استفاده از زبان پایتون و کتابخانه‌های قدرتمندی که دارد پیاده سازی می‌شود و به صورت یک API عمل خواهد کرد که باعث می‌شود وابستگی‌ها به حداقل برسد و قابلیت استفاده مجدد بالا برود. بدین صورت که این مازول‌ها روی port دیگری در حال سرویس دهی خواهند آمد و وب

اپلیکیشن ما هر گاه نیاز به استفاده از آن‌ها داشته باشد با ارسال داده‌های مورد نیاز آن عملکرد با این ماژول‌ها در ارتباط خواهد بود. به منظور داده کاوی بر روی سوالات پرسشنامه‌ها عملیات داده کاوی عبارت خواهد بود از:

- خوشه‌بندی براساس تمامی داده‌ها: این روش به ما کمک خواهد کرد تا افراد بر اساس شباهت نظراتشان دسته بندی شوند. این دسته‌ها در پژوهش‌های بعدی امکان تحلیل تغییرات نظرات یک دسته از جامعه را فراهم خواهد کرد. با توجه به حجم نسبتاً بالای داده‌ها در این بخش از الگوریتم‌های خوشه‌بندی مناسب نظیر الگوریتم‌های خوشه‌بندی DBSCAN افزایشی و یا BFR استفاده خواهیم نمود، تا سرعت و دقت خوشه‌بندی افزایش یابد

- استخراج قوانین وابستگی: (به عنوان مثال قوانینی نظیر اینکه اگر یک فرد دانشجو بوده و از منطقه جغرافیایی خاصی از کشور باشد آنگاه با احتمال مثلاً ۷۰٪ علاقمندی او به برنامه‌های تحلیلی و اخبار است). این گام یک روش مناسب برای یافتن روابط مهم بین متغیرهای موجود در پایگاه داده‌های بزرگ است. برای این منظور از یکی از دقیق‌ترین الگوریتم‌های کاوش قوانین وابستگی استفاده خواهد شد. این الگوریتم از استراتژی جستجوی اول سطح برای شمارش پشتیبان مجموعه گزینه‌ها استفاده می‌کند و با استفاده از یک تابع تولید کاندید، از خصوصیت بستار رو به پایین پشتیبان بهره می‌برد. در این پروژه با استفاده از این الگوریتم قوانین وابستگی بین نظرات را استخراج خواهیم نمود و برای هر قانون میزان حمایت و سطح اطمینان آن نیز ذکر خواهد شد.

- استخراج روابط و همبستگی بین پاسخ‌های سؤالات مختلف (مثلاً رابطه سن افراد و موضوعات مورد اشاره)

- استخراج نظرات دسته یا گروه خاص از افراد جامعه براساس پاسخ به همه سوالات

- علاوه بر موارد ذکر شده عملیات‌های پیش پردازشی نیز پیش از شروع تحلیل‌ها لازم است.

مراحل متن کاوی نیز عبارت خواهد بود از:

با توجه به اینکه نظرات ارائه شده توسط افراد دارای برچسب موضوعی و معنایی نمی‌باشد نمی‌توان از روش‌های یادگیری ماشین با مربی استفاده نمود. در نتیجه در بخش استخراج موضوعات از روش‌های بدون مربی استفاده خواهد شد. برای این منظور الگوریتم‌های LDA و PLSI بهترین گزینه‌ها می‌باشند. از آنجا که ایجاد قابلیت تحلیل برخط داده‌های جدید ورودی یکی از اهداف این پروژه می‌باشد تا بتوان در هر زمان داده جدید را به سیستم اضافه نمود و موضوعات آن را استخراج کرد، به منظور نیل به این هدف، روش LDA مورد استفاده قرار خواهد گرفت. چراکه امکان استنتاج موضوع از داده‌های جدید بخشی از ویژگی‌ها و کارایی‌های این الگوریتم می‌باشد. بسامد کلمات در متن و بسامد معکوس مستندات در مراحل ابتدایی الگوریتم LDA مورد استفاده قرار خواهد گرفت. همچنین محاسبه اطلاعات متقابل کلمات و استخراج کلمات کلیدی نیز در بخش متن کاوی انجام خواهد شد.

یکی از چالش‌های مهم در روش‌های بدون مربی بهره‌گیری از تعداد بالای رکودهای داده یا متن می‌باشد تا بتوان به نتایج قابل قبولی دست یافت. به همین دلیل، همانطور که در بخش قبل ذکر شد، استخراج موضوع در این پروژه در دو مرحله انجام خواهد شد:

- استخراج کل موضوعات از تمام متون
- یافتن ارتباط میان موضوعات و گروه‌های مختلف شرکت کننده. به این صورت، مرحله اول که بخش کلیدی کار است قبل از تفکیک متون براساس نظردهندگان انجام می‌پذیرد تا تمام داده‌ها به‌طور یک‌جا مورد بهره قرار گیرد و نتایج دقیق‌تری حاصل گردد.

در مرحله بعد، پس از استخراج موضوعات اصلی و یافتن ارتباط آن‌ها با گروه‌های نظردهنده، به‌منظور تحلیل افکار مطرح حول هر موضوع از لحاظ مثبت یا منفی بودن، باید از روش‌های کاوش عقاید بهره برد. هدف این بخش پاسخ به این سؤال است که آیا نظرات مطرح‌شده درباره هر موضوع مثبت است و یا بیشتر جنبه منفی و انتقادی دارد. به این منظور می‌توان از الگوریتم‌های مبتنی بر قاعده برای تشخیص نوع نظرات استفاده نمود. شایان ذکر است که بخش عظیمی از نظرات ارسال‌شده در این بخش جنبه پیشنهادی دارد و به‌طور مشخص حاوی نظر مثبت یا منفی حول موضوع مدنظر نمی‌باشد. تفکیک این نوع متون پیشنهادی از متونی که حاوی نظرات و عقاید خاص و مشخص هستند نیز در دستور کار این بخش از پروژه قرار دارد.

در صورت توافق طرفین قرارداد سیستم پیشنهادی قابلیت گسترش و توسعه در موارد زیر را دارد که می‌تواند در یک قرارداد مجزا بعد از توافق طرفین به انجام رسد.

- ارتباط با پایگاه‌های داده موجود در مرکز و استخراج اطلاعات جداول از آن پایگاه‌ها
- برقراری ارتباط میان جداول نظرسنجی‌های مرتبط به هم
- بررسی همبستگی میان پاسخهای سؤالات مختلف در چند پرسشنامه مرتبط به هم
- ارائه خروجی‌های با قابلیت استفاده در مجموعه آفیس
- گذراندن آزمون‌های نرم‌افزاری نظیر آزمون تحمل خرابی، آزمون واسط کاربر و آزمون امنیتی
- آموزش عمومی پرسنل کارفرما برای استفاده از سیستم (آموزش‌های مقدماتی جهت راه‌اندازی و تست سیستم همزمان با تحویل هر فاز انجام خواهد پذیرفت)

**۲.۱.۷. طرح پشتیبانی کاربر نهایی :** تبادل اطلاعات فنی برای کاربر نهایی در این سیستم علاوه بر چند آموزش کوتاه و همچنین داکيومنت شدن توضیحات راجع به هر بخش و در نهایت به صورت tooltip روی هر کامپوننت توضیحی خواهد بود تا کاربران به راحتی با سیستم کار کنند. گرافیک سیستم با استفاده از html, css, jQuery, bootstrap ایجاد می‌شود که منجر به یک گرافیک responsive و داینامیک خواهد بود و چون از کامپوننت‌ها و استایل استاندارد استفاده می‌شود کاربری که به مقدار معمول از سایت‌های معروف استفاده کرده است مشکلی برای استفاده نخواهد داشت و تجربه کاربری خوبی برای وی رقم خواهد خورد. این سایت از چند زبانه بودن به معنای تغییر زبان کل سایت و پنل‌ها پشتیبانی نخواهد کرد. با توجه به اینکه تعامل سیستم با کاربر زیاد خواهد بود و برای راحتی کاربری خطاهایی که مربوط به کار با سیستم است همچون فایل با فرمت اشتباه به صورت پیام‌های خطا به کاربر نمایش داده می‌شود.

۲,۱,۸. طرح جامع پروژه : طرح توسعه و آزمون در بخش خود آورده شده اند.

طرح شکست پروژه همانطور که در بخش ۸-۱ پروپوزال آورده شده است به شکل زیر است :

شماره فعالیت	شرح فعالیت	مدت اجرا (روز)	مدت اجرا کل فاز (روز)	درصد به کل پروژه	پیش نیاز
۱-۱	تهیه گزارش نیازسنجی پروژه (مطابق با بخش ۱ و ۲ پیوست فنی دو برای مستندات تحویل گیری نرم افزار در RFP) (به جز بندهای ۲,۱,۳ و ۲,۱,۴)	۱۵	۱۵	۳۰%	-
۲-۱	سیستم استخراج اطلاعات داده ها و ساخت جداول در پایگاه داده	۲۰	۱۵۰	۳۰%	-
۲-۲	سیستم یکسان سازی فرمت داده ها تبدیل داده ها به فرمت لازم و ساخت پایگاه داده (مطابق با بخش ۱-۲ مشخصات فنی پروژه در RFP) (به جز مبحث Stored Prosedure)	۱۵			
۳-۲	سیستم پیش پردازش پاسخنامه ها (تشخیص نویز و حذف داده های پرت و تصمیم گیری در مورد سوالات بی پاسخ)	۱۵			
۴-۲	سیستم ارائه انواع گزارش های آماری از پاسخ نامه ها (مطابق با بندهای ۱ تا ۲۰ بخش ۱-۴ مشخصات فنی پروژه در RFP)	۴۰			
۵-۲	واسط کاربر ساده بر مبنای پایگاه داده طراحی شده و گزارشات آماری	۳۰			
۶-۲	سیستم مدیریت پرسشنامه ها (مطابق با بندهای ۱ تا ۵ بخش ۱-۳ مشخصات فنی پروژه در RFP) (به جز بند ۶ بخش ۱-۳ مشخصات فنی پروژه در RFP)	۵			



			۵	پیاده‌سازی ماژول تولید خروجی‌های عددی قابل استفاده در اکسل آفیس (مطابق با بند ۷ بخش ۵-۱ و بند ۵ بخش ۶-۱ و بند ۸ بخش ۱-۱ و بندهای ۱ تا ۳ بخش ۸-۱ مشخصات فنی پروژه در RFP) (بجز ویژگی‌های خط و زبان فارسی و خروجی تحت ورد آفیس که در بند ۶ بخش ۵-۱ مشخصات فنی پروژه در RFP ذکر شده)	۷-۲
			۹۰	طراحی و پیاده‌سازی نسخه اولیه سامانه به همراه سطوح دسترسی (مطابق با بندهای ۱ تا ۶ بخش ۱-۱ و بندهای ۱ تا ۷ بخش ۷-۱ مشخصات فنی پروژه در RFP)	۸-۲
			۵	تهیه گزارش فاز ۲	۹-۲
۳	۲۵٪	۱۲۰	۶۰	سیستم داده‌کاوی (مطابق با بندهای ۱ تا ۵ بخش ۵-۱ مشخصات فنی پروژه در RFP)	۱-۳
			۲۰	سیستم پیش پردازش متون نظرات آزاد	۲-۳
			۶۰	سیستم متن کاوی (مطابق با بندهای ۱ تا ۴ بخش ۶-۱ مشخصات فنی پروژه در RFP)	۳-۳
			۶۰	تکمیل واسط کاربر بر اساس فعالیت‌های داده‌کاوی و متن کاوی	۴-۳
			۵	سیستم ذخیره فایل‌های پشتیبان (مطابق با بند ۷ بخش ۱-۱ مشخصات فنی پروژه در RFP)	۵-۳
			۳۰	ارزیابی عملکرد سیستم	۶-۳
			۳۰	تهیه نسخه نهایی و یکپارچه ابزار خروجی (به جز بندها ۸ بخش ۱-۱)	۷-۳

			۵	تهیه گزارش فاز ۳ (مطابق با بخش ۳ و ۴ پیوست فنی دو برای مستندات تحویل‌گیری نرم‌افزار در RFP) (به جز بندهای ۳,۲ و ۴,۱)	۸-۳	
فاز ۳	۱۵ %	۴۵	۱۰	راه‌اندازی سیستم در محل کارفرما	۱-۴	۴
			۱۰	رفع اشکالات عملی کار با سیستم	۲-۴	
			۲۵	ارائه گزارش نهایی (مطابق با بند ۵,۲ بخش ۵ پیوست فنی دو برای مستندات تحویل‌گیری نرم‌افزار در RFP)	۳-۴	

**۲.۱.۹. طرح پشتیبانی :** در یک بازه دو ساله (پس از تحویل نرم افزار) به صورت رایگان و در بازه پنج ساله بعدی با دریافت هزینه نسبت به رفع هر گونه مشکل سیستم نرم افزار موضوع قرارداد اقدام می‌شود. همچنین حسب درخواست مستند کارفرما، خدمات پشتیبانی نرم افزاری مرتبط با نرم افزار کاربردی به صورت راه دور با استفاده از امکانات مخابراتی، حداکثر ظرف مدت ۲۴ ساعت به غیر از ایام تعطیل به کارفرما ارائه می‌شود. در صورت حل نشدن مشکل از راه دور پس از درخواست مستند کارفرما، حداکثر ظرف مدت ۴۸ ساعت به غیر از ایام تعطیل به محل استقرار نرم افزار کاربردی مراجعه و اقدام به رفع اشکال پیش آمده می‌شود. علاوه بر این، درخواست‌های کارفرما برای توسعه سیستم در مدت زمان پشتیبانی بررسی و برآورد هزینه شده و در صورت توافق طرفین اقدام به توسعه می‌شود. در این سیستم تمام فعالیت‌ها شامل ورود و خروج‌ها و تمام دسترسی به داده و تحلیل داده و استفاده از امکانات تحلیلی و محاسباتی سیستم با استفاده از چارچوب لاراول و کتابخانه‌هایی همچون activitylog، گزارش‌گیری به صورت log در پایگاه داده‌ها با جزئیات ذخیره می‌شود و هر زمان که نیاز باشد می‌توان به آن‌ها مراجعه کرد.

**۲.۱.۱۰. طرح مانیتورینگ :** با توجه به تعداد استفاده کنندگان، تعداد درخواست‌های همزمان این سیستم که طبق قرارداد حدود ۲۰ و حجم داده حدود ۱۰ گیگ خواهد بود با سروری با مشخصاتی که پیش از این اعلام شده است، نقاط شکست و بازیابی معنا نخواهد داشت با این وجود این سیستم از کتابخانه larametric برای ثبت تمام خطاها و مشکلات استفاده خواهد کرد و با استفاده از وب سروری همچون nginx یا آپاچی و اضافه شدن سرورها این وب سرور به خوبی load balancing را انجام خواهد داد و درخواست‌ها را بین سرورها پخش خواهد کرد.

**۲.۱.۱۱. طرح عملیات :** تغییرات در سمت‌ها و افراد با حذف آن شخص یا حذف نقش آن شخص و اضافه کردن نقش به فرد جدید به راحتی قابل دستیابی خواهد بود. تغییرات در actionها نیز به دلیل ماژولار بودن کل فرایند بسیار مشخص است. فرایندهایی که لازم است تغییر پیدا کند اگر در قسمت تحلیل داده و متن و محاسبات و گزارش‌های آماری است به دلیل ماژولار بودن و مجزا بودن کفایت یک ماژول جدید و یک endpoint به سرور آن قسمت اضافه شود که زمانی که درخواست به این endpoint داده می‌شود آن ماژول صدا زده شود. همین‌طور برای تغییر در قسمت وب نیز به دلیل استاندارد بودن استایل و همین‌طور معماری MVC چارچوب لاراول به راحتی قابل تغییر است. افراد به صورت عمومی اگر با مرورگرهای معروف همچون گوگل کروم و فایرفاکس کار کرده باشند از نظر مهارتی به مشکلی برخورد خواهند خورد. افراد وارد کننده داده اگر مهارتهایی در سطح ICDL را دارا باشند به راحتی با این سیستم کار خواهند کرد. تحلیلگران برای استفاده از امکانات محاسبات آماری و تحلیل داده و متن کفایت مفاهیم اولیه آمار را بدانند و مدیران نیز مهارتی در حد استفاده از مرورگرهای معروف داشته باشند به راحتی می‌توانند گزارش‌ها را مطالعه کنند و از سیستم استفاده کنند. برای طراحی این سامانه نیاز به نمونه‌های مختلفی است که عبارت است از داده‌های با فرمت‌های CSV، SAV، schema و درجداول SQL و چندین سطر از جداول آن برای آشنایی با ساختار داده‌ها، چند نمونه از تحلیل‌هایی که تحلیلگران به صورت دستی انجام می‌دادند، متن‌های نظرسنجی به اندازه کافی برای تحلیل‌های متن و نمونه گذشته آن‌ها توسط اپراتورها. همچنین برای آشنایی با سلیقه و عادات کاربری و شبکه داخلی سازمان و سطوح دسترسی، نیاز است با کارشناسان تحلیلگر و مدیران استفاده کننده از تحلیل صحبت شود و در ادامه نیز صحبت با آن‌ها در فازهای مختلف و دریافت پیشنهادهای تجارب کاربری آن‌ها در استفاده از سیستم جدید لازم، ضروری و بهبود دهنده کیفیت خواهد بود.

**۲.۱.۱۲. طرح کارایی :** زمان پاسخگویی برای هر عملکرد نیازمند دانستن میزان ظرفیت شبکه‌ها و سرور سازمان است اما برای عملکردهای وب اپلیکیشن و خواندن داده از روی پایگاه‌های داده زمان در نظر گرفته شده کمتر از یک ثانیه است که کاربر به هیچ وجه تاخیری را حس نکنند ولی زمان تحلیل داده و متن وابسته به داده دریافتی است ولی در مجموع باز هم زمان آن در حدی نخواهد بود که تاخیر محسوسی در سیستم به چشم بخورد.

**۲.۱.۱۳. طرح پایلوت :** فاز دوم این سیستم به صورت پایلوت روی سرورهای سازمان به اجرا در می‌آید تا بازخوردها و تجربه‌های کاربری افراد مختلف گرفته شده و در فازهای بعدی به کار گرفته شود و سیستم بهبود پیدا کند. این فاز شامل سیستم یکسان‌سازی فرمت داده‌ها تبدیل داده‌ها به فرمت لازم و ساخت پایگاه داده، سیستم پیش‌پردازش پاسخنامه‌ها (تشخیص نویز و حذف داده‌های پرت و تصمیم‌گیری در مورد سوالات بی‌پاسخ)، سیستم ارائه انواع گزارش‌های آماری از پاسخ نامه‌ها، واسط کاربر ساده بر مبنای پایگاه داده طراحی شده و گزارشات آماری، سیستم مدیریت پرسشنامه‌ها، پیاده‌سازی ماژول تولید خروجی‌های عددی قابل استفاده در اکسل آفیس و طراحی و پیاده‌سازی نسخه اولیه سامانه به همراه سطوح دسترسی است. در این فاز هدف بررسی راحتی کاربری سیستم، پیش‌پردازش صحیح داده‌ها و تکمیل داده‌های پاسخ داده نشده، ساخت پایگاه داده‌ها، امنیت لازم در دسترسی کاربران تنها به داده‌های سطح خود و گزارشات آماری صحیح برای داده هاست.

**۲.۱.۱۴. طرح امنیت :** در این سیستم سعی بر این است که حداکثر امنیت ممکن از هر سه منظر یکپارچگی، محرمانگی و در دسترس بودن تحقق یابد. با توجه به سیستم backup and restore همواره داده‌ها چندباره ذخیره می‌شوند و همواره در دسترس بودن داده‌ها تامین می‌شود. با توجه به اینکه استانداردهای چک لیستی همچون چک لیست OWASP برای تست یک وب اپلیکیشن وجود دارد، در صورت رعایت این نکات در جزئیات پیاده سازی اکثر حفره‌های امنیتی و مشکلات بهبود پیدا می‌کنند. از این رو ما خود را ملزم می‌دانیم که طبق چک لیست تمام آستانه تحمل‌ها و متدهای خود را پیش ببریم. بعضی از حملات با گذاشتن قوانین خنثی می‌شوند. که از آن جمله می‌توان به ملزم ساختن کاربر به رمزهای قوی و متشکل از حروف و اعداد و علائم و حداقل هشت حرفی گذاشتن و رعایت نکات اشاره کرد. اما بعضی از حملات نیز با توجه به امکاناتی که خود سازمان برای این سیستم در نظر گرفته است بهبود پیدا می‌کند برای مثال برای جلوگیری از حمله DDOS (منع کاربری توزیع شده) برنامه نویسی کار خاصی را نمی‌تواند انجام دهد و کارفرما با توجه به صلاحدید می‌بایست سرویس خود را روی سرورهای ابری شرکت‌های ارائه دهنده خدمات CDN ببرد البته برای یک سیستمی که قرار نیست به صورت عمومی استفاده شود DDOS حمله نادری خواهد بود اما با توجه به اسکنرهایی که هکرها روی IPهای سازمان قرار می‌دهند همچنان خطر وجود دارد. در چارچوب شناخته شده‌ای مانند لاراول در صورت استفاده نکردن از کتابخانه‌های تایید نشده که این مورد حتما رعایت خواهد شد جلوی اکثر حملات همچون sql injection توسط این چارچوب گرفته می‌شود اما pen test نهایی نیز بعد از اتمام کار می‌تواند باعث بهبود هر چه بیشتر امنیت سیستم شود. با توجه به میزان محرمانگی و اهمیت امنیت برای سازمان نیز می‌توان تست‌هایی همچون red team را نیز به این فرآیند اضافه کرد که می‌بایست این تست‌ها به صورت جداگانه در قرارداد بعدی آورده شوند.

**۲.۱.۱۶. طرح آزمون :** سناریوهای تست مهم سامانه با استفاده از چارچوب تست selenium نوشته می‌شود و کاربری‌های مختلف سامانه تست خواهد شد. عملکردهایی همچون اضافه کردن داده، محاسبات و گزارش‌های آماری مختلف و تحلیل داده و متن و گزارش‌گیری تست خواهد شد. همچنین بعد از انتهای هر فاز با تست‌های کاربری که کارشناسان و مدیران انجام می‌دهند تست دوباره‌ای روی آن صورت می‌گیرد و سیستم بهبود داده می‌شود. برای گزارش باگ‌ها و مشکلات از سرویس جامع hotjar استفاده می‌شود که به صورت بصری و بسیار راحت امکان گزارش مشکلات را فراهم می‌کند و باعث بهبود هر چه بهتر سیستم می‌شود.

## ۲.۲. مشخصات

**۲.۲.۱. نیازمندی‌های کسب و کار :** در این پروژه گزارشات آماری و تحلیل جامعی از داده‌های متنی و چندگزینیه‌ای و ارتباط بین پرسشنامه‌ها ارائه می‌شود که منجر به یکپارچه‌سازی و مجتمع‌سازی داده‌ها و تحلیل‌ها در یک پایگاه داده و سیستم می‌شود که در آینده نیز می‌توان به یکپارچه‌سازی کل سیستم از ابتدای واردکردن داده‌ها توسط اپراتورها و در ادامه یکپارچه‌سازی به وارد کردن مستقیم داده‌ها توسط پرسش‌کنندگان میدانی، اتصال به پایگاه‌های اینترنتی که نظرسنجی در آن‌ها انجام می‌شود و سیستم پیامکی سازمان روی آورد. برای طراحی این سامانه نیاز به نمونه‌های مختلفی است که عبارت است از داده‌های با فرمت‌های CSV، SAV و schema در جداول SQL و چندین سطر از جداول آن برای آشنایی با ساختار داده‌ها، چند نمونه از تحلیل‌هایی که تحلیلگران به صورت دستی انجام می‌دادند، متن‌های نظرسنجی به اندازه کافی برای تحلیل‌های متن و نمونه کد شده آن‌ها توسط اپراتورها. همچنین برای آشنایی با سلیقه و عادات کاربری و شبکه داخلی سازمان و سطوح دسترسی، نیاز است با کارشناسان تحلیلگر و مدیران استفاده‌کننده از تحلیل صحبت شود و در ادامه نیز صحبت با آن‌ها در فازهای مختلف و دریافت پیشنهادهای تجارب کاربری آن‌ها در استفاده از سیستم جدید لازم، ضروری و بهبود دهنده کیفیت خواهد بود. امنیت این سیستم به علت رعایت استانداردهای کدزنی و چک لیست‌های امنیتی و تست‌های لازم بسیار مطلوب خواهد بود که در بخش طرح امنیت توضیح داده شد. افرادی که با این سیستم کار می‌کنند نیازمند داشتن دانش سطح بالا در علوم کامپیوتر و علم داده نیستند بلکه به دلیل محیط کاربری خوبی که این سیستم داراست به صورت کلی همه کسانی که با سایت‌ها و مرورگرهای معروف کار کرده‌اند به راحتی از این سیستم استفاده می‌کنند و تحلیلگران نیز با دانستن مفاهیم اولیه آمار به راحتی نیازهای تحلیلی خود را برآورده می‌سازند.

**۲.۲.۲. مشخصات عملکردی :** در این پروژه گزارشات آماری و تحلیل جامعی از داده‌های متنی و چندگزینیه‌ای و ارتباط بین پرسشنامه‌ها ارائه می‌شود که منجر به یکپارچه‌سازی و مجتمع‌سازی داده‌ها و تحلیل‌ها در یک پایگاه داده و سیستم می‌شود که در آینده نیز می‌توان به یکپارچه‌سازی کل سیستم از ابتدای واردکردن داده‌ها توسط اپراتورها و در ادامه یکپارچه‌سازی وارد کردن مستقیم داده‌ها توسط پرسش‌کنندگان میدانی، اتصال به پایگاه‌های اینترنتی که نظرسنجی در آن‌ها انجام می‌شود و سیستم پیامکی سازمان روی آورد. ویژگی‌ها و عملکردهای سیستم به صورت کلی در قالب یک usecase دیاگرام در زیر آورده شده است که نقش‌ها و عملکرد کلی هر یک قابل مشاهده هستند. توضیحات هر عملکرد در بخش حوزه چشم اندازه به خوبی آورده شده است. در اینجا اهداف و پیش فرض‌ها را برای هر یک از usecase خواهیم داشت :

- واردکردن داده (import data) : هدف واردشدن داده‌های صورت سوال‌های پرسشنامه به همراه تمامی پاسخ‌ها اعم از متنی و غیرمتنی می‌باشد تا یک سیستم یکپارچه از سوال‌ها و پاسخ‌نامه‌ها برای تحلیل داشته باشیم. پیش فرض این بخش این است که اپراتورها باید سطح دسترسی لازم (دسترسی به پرسش‌نامه‌ها و دسترسی واردکردن سوال‌ها و پاسخ‌نامه‌ها) را از ادمین دریافت کرده و به حساب خود وارد شده باشند.

- مدیریت پرسشنامه‌ها (manage questionnaire) : هدف مدیریت پرسش‌نامه‌ها عبارت است از مشاهده آمار کلی مانند تعداد پرسش‌نامه‌ها و پاسخ‌ها، دسته بندی، قفل و حذف آن‌ها و همچنین یکسان سازی پارامترهای مختلف در پرسشنامه‌ها. پیش‌فرض این بخش این است که کاربر باید سطح دسترسی لازم (سطح دسترسی برای نقش تحلیلگر) برای مدیریت پرسشنامه‌ها را داشته و به حساب خود وارد شده باشد.
- گزارش‌ها و محاسبات آماری (get statistical report) : هدف انجام مشخصه‌ها و محاسبات آماری همچون ضرایب همبستگی و آزمون‌های آماری با استفاده از رابط کاربری است که شرح مفصل جزئیات در بخش حوزه چشم انداز آورده شده است. پیش‌فرض این بخش این است که کاربر باید سطح دسترسی لازم (سطح دسترسی برای نقش تحلیلگر) را داشته و به حساب خود وارد شده باشد.
- یافتن الگوهای پرتکرار (get repetitive pattern) : هدف یافتن الگوهایی همچون تفاوت میزان مشارکت در انتخابات درمورد طرفداران یک کاندیدا در پاسخ‌هاست که این الگوهای پنهان به صورت گسترده در بین مردم یا قشر خاصی از مردم وجود دارد که با تحلیل قوانین انجمنی بدست می‌آیند و می‌تواند به تحلیلگر در شناخت مردم و مخاطبین کمک زیادی کند. تحلیلگر مانند قسمت‌های قبل باید سطح دسترسی لازم را دارا باشد و به حساب خود وارد شده باشد.
- یافتن گروه‌های مشابه (get similar groups) : هدف یافتن گروه‌های مختلف مردم با توجه به پاسخ‌هایی است که به پرسشنامه داده‌اند که باعث دسته بندی مردم و راحت‌تر شدن تصمیم‌گیری‌های برای هر دسته خواهد شد. کاربر باید سطح دسترسی لازم (سطح دسترسی برای نقش تحلیلگر) را داشته و به حساب خود وارد شده باشد. شایان ذکر است که جزییات روش مورد استفاده در تمام فعالیت‌های مربوط به داده‌کاوی و متن‌کاوی در طرح مفهومی ذکر گردیده‌است.
- پیش‌بینی رفتار کاربران (predict behaviors) : هدف این بخش پیش‌بینی رفتار مخاطبان می‌باشد. با توجه به دسته بندی کاربران و همچنین یافتن الگوهای تکراری ارائه پیش‌بینی رفتارها و پاسخ‌های کاربران بسیار راحت‌تر خواهد بود که موجب می‌شود برنامه ریزی و جهت گیری در قبال آن‌ها بسیار ساده‌تر شود. کاربر باید سطح دسترسی لازم (سطح دسترسی برای نقش تحلیلگر) را داشته و به حساب خود وارد شده باشد.
- خروجی گرفتن از پرسشنامه (export questionnaire) : هدف این بخش این است که پاسخنامه‌هایی که باهم تجمیع شده‌اند در قالب سؤالات ادغام‌شده به کاربر نمایش داده‌شوند به‌صورتی که پرسشنامه هر رکورد در فایل تجمیعی مشخص شده باشد.
- پیش‌پردازش داده‌ها و متن‌ها (preprocess data and text) : هدف حذف داده‌های پرت، جانمایی پرسش‌های بی پاسخ، ارتباط بین پرسشنامه‌های مختلف با فیلدهای یکسان<sup>۱</sup>، تبدیل متون عامی به صورت رسمی می‌باشد. این امر باعث تشخیص راحت‌تر مرز میان واژه‌ها و یکسان سازی کلمات و تصمیم‌گیری در قبال سوال‌های بدون پاسخ است.
- تحلیل افکار در پرسش‌های آزاد (analyse thoughts in free questions) : پس از استخراج موضوعات اصلی و یافتن ارتباط آنها با گروه‌های نظردهنده، به منظور تحلیل افکار مطرح حول هر موضوع از لحاظ مثبت

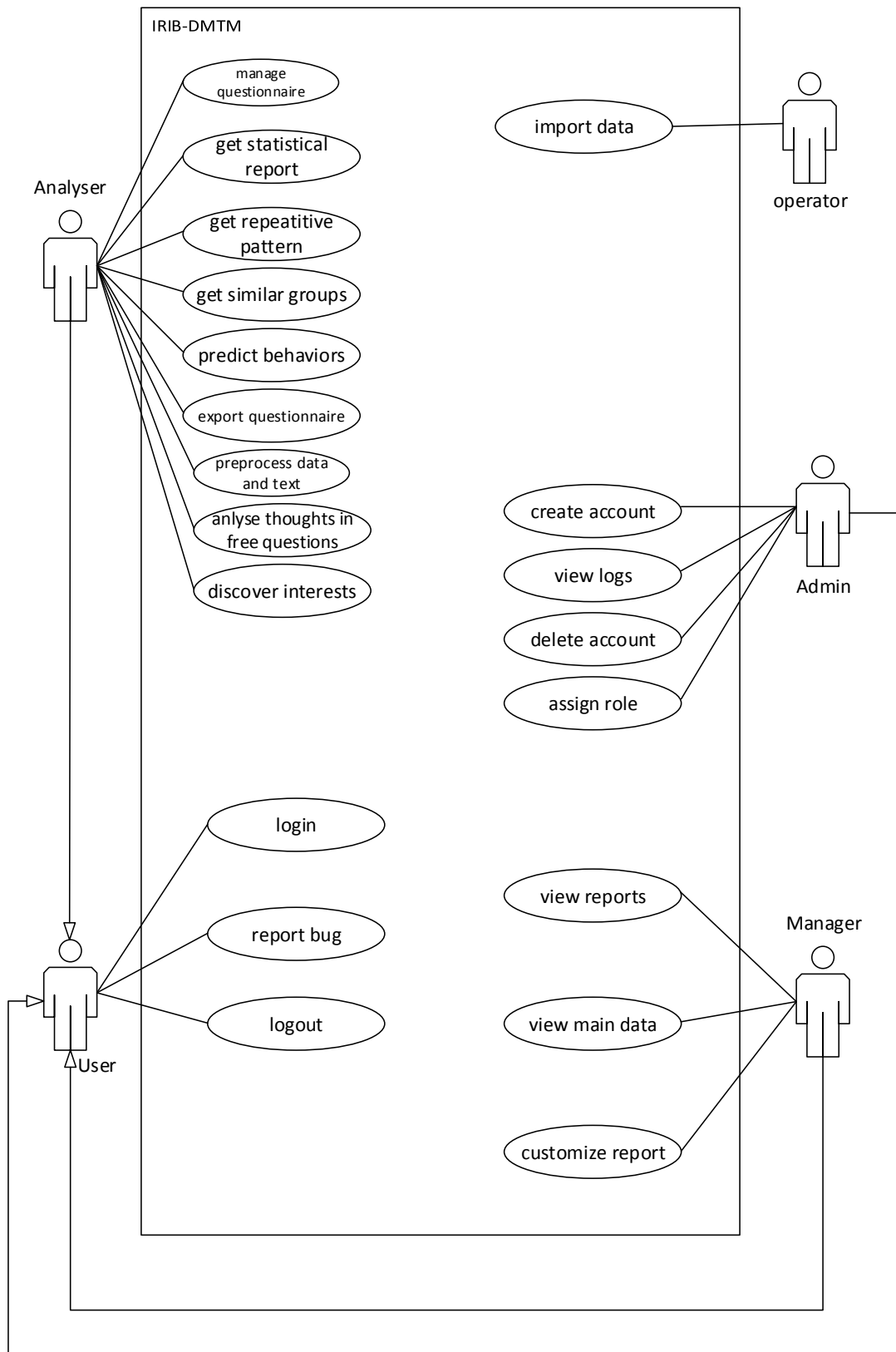
<sup>۱</sup> بخش تحلیل روابط بین پرسشنامه‌ها در پروژه گسترش یافته اضافه خواهد شد

یا منفی بودن، باید از روش‌های کاوش عقاید بهره برد. هدف این بخش پاسخ به این سؤال است که موضوعات اصلی مطرح شده در متون باز چه مواردی هستند و آیا نظرات مطرح شده درباره هر موضوع مثبت است و یا بیشتر جنبه منفی و انتقادی دارد.

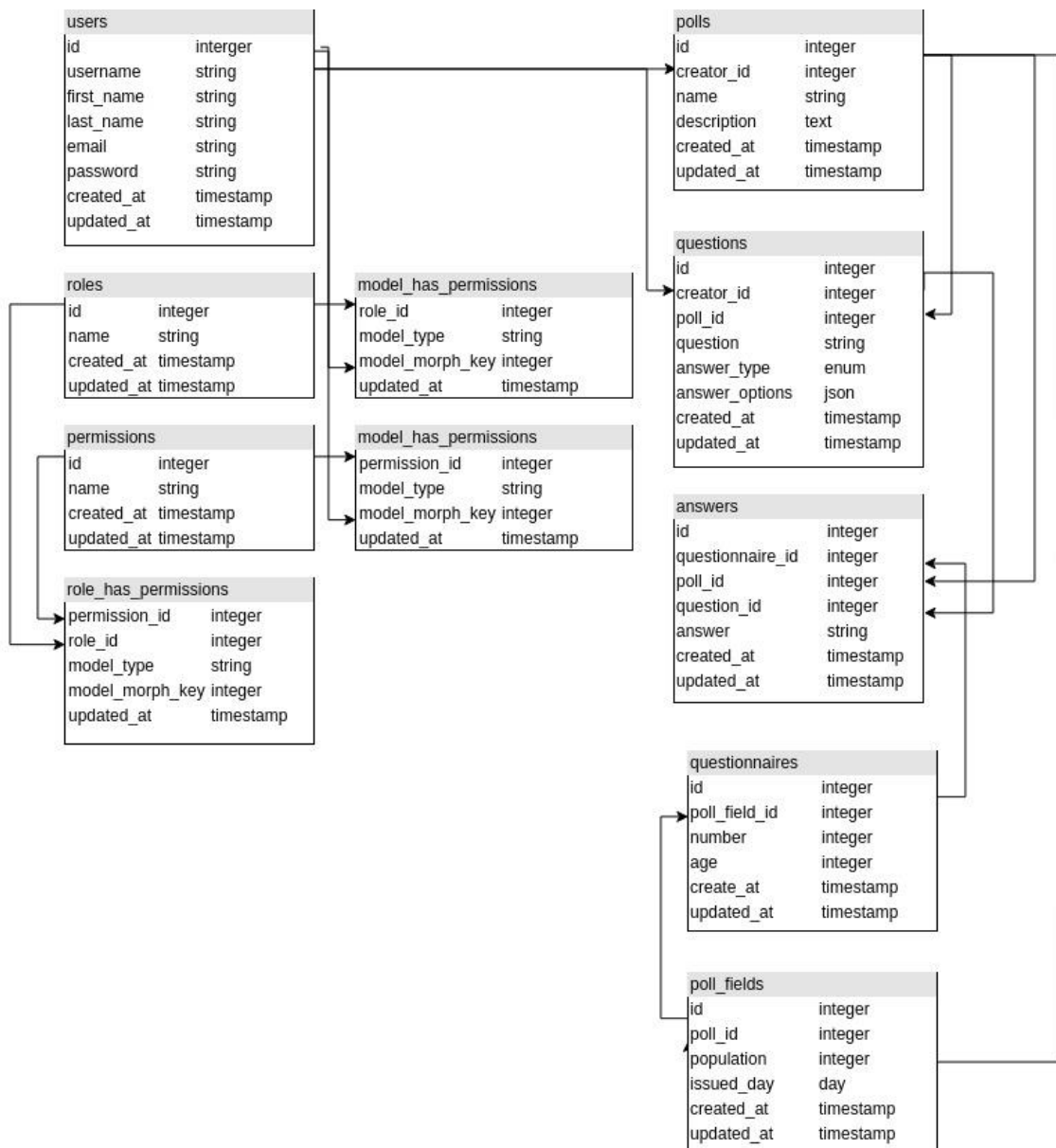
- کشف علاقه‌مندی‌ها (discover interests) : یافتن علایق و نظرات گروه‌های مختلف شرکت‌کننده در نظرسنجی بر مبنای تفکیک رده سنی، جنسیت، میزان تحصیلات و محل سکونت و کل ویژگی‌های موجود در نظرسنجی‌ها هدف این قسمت است و کاربر می‌تواند در این بخش یکی از ویژگی‌های داده‌ها را مشخص کند و متون باز را بر اساس آن ویژگی فیلتر نماید.
- ساختن حساب کاربری (create account) : هدف این بخش این است که ادمین بتواند برای کاربرانی که از این سیستم استفاده می‌کنند حساب کاربری با استفاده از پنل ادمین ایجاد نماید و در اختیار کاربر قرار دهد.
- مشاهده ریزگزارش‌ها (view logs) : هدف این بخش این است که ادمین بتواند گزارش‌های اتفاقات و عملکردهایی که در سیستم ثبت می‌شود را به صورت کامل مشاهده کند و در فواصل تاریخی مورد نیاز گزارش را فیلتر کند
- حذف حساب کاربری (delete account) : هدف این است که ادمین با حذف کاربر از پایگاه داده کاربران با استفاده از پنل مدیریتی بتواند استفاده کاربر از سیستم را مسدود کند.
- اختصاص نقش (assign role) : هدف این است که ادمین با نقش‌هایی که به افراد اختصاص می‌دهد دسترسی آن‌ها به قسمت‌های مختلف سامانه و اختیارات آن‌ها را مشخص و محدود کند. این نقش‌ها قابل تغییر و حذف هستند.
- ورود به حساب کاربری (login) : هدف این است که همه کاربران با استفاده از نام کاربری و پسوردی که مخصوص به خود دارند بتوانند در سامانه وارد شوند و از سیستم استفاده کنند. کاربر قادر به بازیابی پسورد خود در صورت فراموشی با استفاده از ایمیل سازمانی خود است.
- گزارش باگ و اشکال (report bug) : همه کاربران سیستم می‌توانند مشکلاتی که در سیستم مشاهده می‌کنند گزارش کنند که این مشکلات به دست تیم فنی پروژه می‌رسد و راجع به آن‌ها تصمیم‌گیری می‌شود و در صورت لزوم مشکل رفع خواهد شد. هدف این بخش فراهم نمودن امکان گزارش این مشکلات و پیگیری نتیجه می‌باشد.
- خروج از حساب کاربری (logout) : هدف این بخش این است که کاربر بتواند خودش را از سیستم خارج کند. همچنین در صورت استفاده نکردن از سیستم به مدتی از سیستم به صورت خودکار خارج خواهد شد.
- مشاهده گزارش‌ها (view reports) : هدف این بخش این است که مدیران بتوانند گزارش‌هایی که توسط تحلیلگران ساخته شده را به صورت دسته بندی شده مشاهده نمایند. این گزارش‌ها به صورت دسته بندی شده هستند و قابل شخصی سازی هستند.
- مشاهده داده‌های اصلی (view main data) : هدف این بخش فراهم نمودن این امکان است که مدیران بتوانند علاوه بر گزارش‌ها داده‌های اصلی که از روی آن‌ها گزارش و تحلیل ارائه شده را مشاهده نمایند.
- شخصی سازی قالب گزارش (customize report) : مدیرانی که گزارش‌ها در اختیار آن‌ها قرار می‌گیرد می‌توانند قالب گزارش‌ها را ویرایش و به شکل دلخواه در بیاورند. هدف این بخش فراهم نمودن این امکان برای

مدیران می‌باشد. به‌صورتی که بتوانند تغییراتی مانند رسم انواع نمودارهای متنوع آماری Pie، میله ای و.. با استفاده از رابط کاربری، تغییر رنگ و فونت و اندازه گزارش را در سیستم ایجاد نمایند. این گزارش‌ها با توجه به responsive بودن وب اپلیکیشن در گوشی نیز به راحتی قابل نمایش است که باعث می‌شود مدیران به راحتی در هر زمانی به نتایج دسترسی پیدا کنند.





**۲.۲.۳. طرح منطقی :** نمودار استاندارد روابط بین موجودیت‌ها که نشان دهنده یک طرح اولیه مشخص برای پایگاه داده سیستم و طراحی موجودیت‌ها و مشخصات و رفتارهای آن‌ها خواهد بود. این نمودار ممکن است در ابتدای فاز دو دستخوش تغییراتی نیز شود.



**۲.۲.۴. طرح مفهومی :** معماری راه حل به طور کامل در بخش طرح توسعه توضیح داده شد. با توجه به usecase diagram که در بخش مشخصات عملکردی آورده شده است برای هر کدام از آن‌ها خواهیم داشت :

- مدیریت پرسشنامه‌ها : کاربر بعد از ورود در سیستم با توجه به سطح دسترسی که توسط ادمین برای وی تعریف شده است به پرسشنامه‌های سطح خود دسترسی دارد، می‌تواند با رابط کاربری مناسب آمار کلی و تعداد

- پاسخ آن‌ها را ببیند، آن‌ها را دسته بندی، قفل، حذف و ویرایش کند و پارامترهای مشابه را در بین نظرسنجی‌های مختلف یکسان کند.
- گرفتن گزارشات آماری : تحلیلگر پس از ورود می‌تواند تحلیل آماری را با استفاده از وب اپلیکیشن به راحتی و با استفاده از فشردن چند دکمه و با انتخاب پرسشنامه‌های لازم انجام دهد که شرح گزارش‌ها و محاسبات آماری در بخش حوزه چشم انداز آورده شده است، این وب اپلیکیشن برای محاسبه آن‌ها درخواستی به مازول‌های محاسباتی می‌فرستد که شرح آن در بخش توسعه آمد. همچنین کاربر با استفاده از چندین روش پرسش‌های بی پاسخ را پر می‌کند.
  - یافتن الگوهای پرتکرار، یافتن گروه‌های مشابه، پیش بینی رفتار : هر یک از این سناریوهای استفاده بخشی از فرایند زیر خواهد بود :
    - خوشه‌بندی براساس تمامی داده‌ها: این روش به ما کمک خواهد کرد تا افراد بر اساس شباهت نظراتشان دسته بندی شوند. این دسته‌ها در پژوهش‌های بعدی امکان تحلیل تغییرات نظرات یک دسته از جامعه را فراهم خواهد کرد. با توجه به حجم نسبتاً بالای داده‌ها در این بخش از الگوریتم‌های خوشه بندی مناسب نظیر الگوریتم‌های خوشه بندی DBSCAN افزایشی و یا BFR استفاده خواهیم نمود، تا سرعت و دقت خوشه بندی افزایش یابد.
    - استخراج قوانین وابستگی: (به‌عنوان مثال قوانینی نظیر اینکه اگر یک فرد دانشجو بوده و از منطقه جغرافیایی خاصی از کشور باشد آنگاه با احتمال مثلاً ۷۰٪ علاقمندی او به برنامه‌های تحلیلی و اخبار است.) این گام یک روش مناسب برای یافتن روابط مهم بین متغیرهای موجود در پایگاه داده‌های بزرگ است. برای این منظور از یکی از دقیق‌ترین الگوریتم‌های کاوش قوانین وابستگی استفاده خواهد شد. این الگوریتم از استراتژی جستجوی اول-سطح برای شمارش پشتیبان مجموعه گزینه‌ها استفاده می‌کند و با استفاده از یک تابع تولید کاندید، از خصوصیت بستار رو به پایین پشتیبان بهره می‌برد. در این پروژه با استفاده از این الگوریتم قوانین وابستگی بین نظرات را استخراج خواهیم نمود و برای هر قانون میزان حمایت و سطح اطمینان آن نیز ذکر خواهد شد.
    - استخراج روابط و همبستگی بین پاسخ‌های سؤالات مختلف (مثلاً رابطه سن افراد و موضوعات مورد اشاره)
    - استخراج نظرات دسته یا گروه خاص از افراد جامعه براساس پاسخ به همه سؤالات
  - پیش پردازش داده : پیش پردازش داده‌ها شامل حذف داده‌های پرت، جانپی پرسش‌های بی پاسخ (روش‌ها در بخش گزارش حوزه چشم‌انداز آورده شده‌اند) و ارتباط بین پرسشنامه‌های مختلف با فیلدهای یکسان است که در بخش گسترش یافته پروژه اضافه می‌شود.
  - پیش پردازش متن : در این بخش متونی که توسط افراد عامی نوشته شده‌اند به صورت رسمی تبدیل می‌شود تا تشخیص مرز میان واژه‌ها و یکسان سازی کلمات با حالت‌ها راحت‌تر انجام گیرد. همچنین داده‌هایی که به صورت دیجیتالی جمع شده است و مشکلاتی همچون نوشتن فارسی با حروف لاتین، تغییر در کلمات به دلیل نبود بعضی از حروف در کیبورد رفع شود.

- تحلیل افکار در پرسش‌های آزاد و یافتن علاقه‌مندی‌ها: این دو سناریوی استفاده نیز با روش زیر بدست خواهند آمد: با توجه به اینکه نظرات ارائه‌شده توسط افراد دارای برچسب موضوعی و معنایی نمی‌باشد نمی‌توان از روش‌های یادگیری ماشین با مربی استفاده نمود. در نتیجه در بخش استخراج موضوعات از روش‌های بدون مربی استفاده خواهد شد. برای این منظور الگوریتم‌های LDA و PLSI بهترین گزینه‌ها می‌باشند. از آنجا که ایجاد قابلیت تحلیل برخط داده‌های جدید ورودی یکی از اهداف این پروژه می‌باشد تا بتوان در هر زمان داده جدید را به سیستم اضافه نمود و موضوعات آن را استخراج کرد، به‌منظور نیل به این هدف، روش LDA مورد استفاده قرار خواهد گرفت. چراکه امکان استنتاج موضوع از داده‌های جدید بخشی از ویژگی‌ها و کارایی‌های این الگوریتم می‌باشد. بسامد کلمات در متن و بسامد معکوس مستندات در مراحل ابتدایی الگوریتم LDA مورد استفاده قرار خواهد گرفت. همچنین محاسبه اطلاعات متقابل کلمات و استخراج کلمات کلید نیز در بخش متن‌کاوی انجام خواهد شد. یکی از چالش‌های مهم در روش‌های بدون مربی بهره‌گیری از تعداد بالای رکودهای داده یا متن می‌باشد تا بتوان به نتایج قابل قبولی دست‌یافت. به‌همین دلیل، همانطور که در بخش قبل ذکر شد، استخراج موضوع در این پروژه در دو مرحله انجام خواهد شد: (۱) استخراج کل موضوعات از تمام متون (۲) یافتن ارتباط میان موضوعات و گروه‌های مختلف شرکت‌کننده. به این صورت، مرحله اول که بخش کلیدی کار است قبل از تفکیک متون براساس نظردهندگان انجام می‌پذیرد تا تمام داده‌ها به‌طور یک‌جا مورد بهره قرار گیرد و نتایج دقیق‌تری حاصل گردد. در مرحله بعد، پس از استخراج موضوعات اصلی و یافتن ارتباط آن‌ها با گروه‌های نظردهنده، به‌منظور تحلیل افکار مطرح حول هر موضوع از لحاظ مثبت یا منفی بودن، باید از روش‌های کاوش عقاید بهره برد. هدف این بخش پاسخ به این سؤال است که آیا نظرات مطرح‌شده درباره هر موضوع مثبت است و یا بیشتر جنبه منفی و انتقادی دارد. به این منظور می‌توان از الگوریتم‌های مبتنی بر قاعده برای تشخیص نوع نظرات استفاده نمود. به‌علاوه، از آنجا که در بخش استخراج موضوعات روش LDA مورد استفاده قرار خواهد گرفت، در این بخش نیز می‌توان از مدل‌های جدید کاوش اعتقادات که بر مبنای همین روش عمل می‌کنند بهره برد به‌طوری‌که استخراج موضوع و نوع نظر مطرح‌شده حول هر موضوع به‌صورت ترکیبی انجام پذیرد تا بتوان به یک تحلیل یکپارچه در این زمینه دست یافت. شایان ذکر است که بخش عظیمی از نظرات ارسال‌شده در این بخش جنبه پیشنهادی دارد و به‌طور مشخص حاوی نظر مثبت یا منفی حول موضوع مدنظر نمی‌باشد. تفکیک این نوع متون پیشنهادی از متونی که حاوی نظرات و عقاید خاص و مشخص هستند نیز در دستور کار این بخش از پروژه قرار دارد.
- شخصی‌سازی داشبورد مدیریتی: شخصی که دسترسی مدیریتی برای وی تعریف شده است می‌تواند با استفاده از داشبورد مدیریتی تمام گزارشات ایجاد شده در بخش‌های قبل با توجه به نوع دسترسی را ببیند و به اطلاعات اصلی پرسشنامه‌ها دسترسی پیدا کند انواع نمودارهای متنوع آماری Pie، میله ای و.. را با استفاده از رابط کاربری به راحتی رسم کند و گزارشات را هم به صورت جدولی و هم نموداری مشاهده کند. همچنین رنگ و فونت و اندازه گزارش با سلیقه وی تغییر کند. این گزارشات با توجه به responsive بودن وب اپلیکیشن در گوشی نیز به راحتی قابل نمایش است که باعث می‌شود مدیران به راحتی در هر زمانی به نتایج دسترسی پیدا کنند.

**۲,۲,۵. نیازمندی‌های عملیاتی :** علاوه بر نیازهایی که در بخش طرح پشتیبانی به صورت کامل آورده شده است، برای طراحی این سیستم به دو کامپیوتر شخصی پیاده سازی نیاز است که در دانشگاه این مورد محقق می‌شود و در انتهای هر فاز نیاز است به سرور مورد نظر در سازمان صداسیما و همچنین یکی از کامپیوترهای شخصی که توسط کاربران به صورت معمول از آن استفاده می‌شود دسترسی پیدا کنیم تا استقرار و تست لازم انجام شود. نکات امنیتی این سیستم نیز در بخش طرح امنیت کاملاً توضیح داده شده است.

**۲,۲,۶. طرح فیزیکی :** این سیستم در مرورگر کاربر اجرا می‌شود پس به سیستم عامل وی بستگی ندارد و کاربر با یک کامپیوتر با مشخصات متوسط به راحتی می‌تواند از سیستم استفاده کند. سرور و مشخصات مورد نیاز برای این سیستم قبلاً ارائه شده است اما به هیچوجه این سیستم به برند سرور یا سیستم عامل آن محدود نیست. معماری و مشخصات دقیق کامپوننت‌ها، رفتار و ارتباط بین آن‌ها و شرح دقیق سرویس‌های تحلیل داده و متن به صورت دقیق در طرح توسعه و مشخصات امنیت و استقرار آن نیز به ترتیب در طرح امنیت و استقرار کاملاً توضیح داده شده است.

**۲,۲,۸. سناریوهای استفاده :** نحوه اجرای هر کدام از سناریوهای استفاده در بخش طراحی مفهومی به صورت مفصل آورده شده است.

**۲,۲,۹. نیازمندی‌های کاربر :** با توجه به اینکه تا قبل از این کاربران تجربه کاربری با سیستم مشابه را ندارند و تنها محاسبه متغیرهای آماری به کمک SPSS انجام می‌دادند در این سیستم سعی شده است طراحی وب اپلیکیشن مطابق طراحی‌های استاندارد باشد و سادگی طراحی در عین زیبایی در همه جا رعایت شود تا افرادی که تجربه کاربری در سایت‌های معروف را دارا هستند به راحتی با سیستم کار کنند و در استفاده از ماژول‌ها به مشکلی برخوردند و البته با جمع‌آوری بازخوردها در فازهای مختلف این سیستم بهبود خواهد یافت. با توجه به اینکه این سیستم چندین مرتبه با توجه به توضیحاتی که در بخش طرح تست توضیح داده شد، تست خواهد شد و همچنین اشکالات وارده همانطور که در بخش طرح پشتیبانی توضیح داده شد به خوبی جمع‌آوری می‌شود و از نظر امنیت نیز همانطور که در بخش طرح امنیت توضیح داده شد در سطح بسیار خوبی باشد انتظار می‌رود این سیستم کاملاً قابل اعتماد و در دسترس باشد و هیچگونه خللی در استفاده از آن پیش نیاید. این سیستم فارغ از سیستم عامل و شرکت سازنده سرور و کاربر نهایی در مرورگر کاربر اجرا می‌شود و تنها با به‌روز بودن مرورگر توسط کاربر به راحتی قابل استفاده است. افراد به صورت عمومی اگر با مرورگرهای معروف همچون گوگل کروم و فایرفاکس کار کرده باشند از نظر مهارتی به مشکلی برخورد نخواهند. افراد وارد کننده داده اگر مهارت‌هایی در سطح ICDL را دارا باشند به راحتی با این سیستم کار خواهند کرد. تحلیلگران برای استفاده از امکانات محاسبات آماری و تحلیل داده و متن کفایت مفاهیم اولیه آمار را بدانند و مدیران نیز مهارتی در حد استفاده از مرورگرهای معروف داشته باشند به راحتی می‌توانند گزارش‌ها را مطالعه کنند و از سیستم استفاده کنند. این وب اپلیکیشن از چند زبانه بودن پشتیبانی نخواهد کرد.