## شرح اهداف پروژه

هدف اصلی این پروژه ارائه سیستمی برای تحلیل خودکار نظرات افرادی است که در نظر سنجی­های صدا وسیما شرکت نموده‌اند و به سؤالات چند گزینه‌ای و همین‌طور سوال پیشنهاد و نظر آزاد پاسخ داده‌اند. در این تحلیل ارتباط بین پاسخ­های افراد به سوالات 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 به‌تبع آن ارائه راهکارهای مفید در رفع کاستی­ها میسرتر خواهد بود.

## دستاوردهای طرح

خروجی‌ اصلی این طرح یک ابزار یکپارچه برای تحلیل داده‌ها و متون نظرسنجی است. نسخه اول این سامانه قادر خواهد بود به ازای دریافت اطلاعات بندهای زیر را به انجام رساند:

* تبدیل فرمت فایل ورودی به جداول پایگاه داده
  + دریافت فایل داده­های نظر سنجی به فرمت Sav و یا سایر فرمت­هایی که در قرار داد ذکر خواهد شد.
  + دریافت عناوین سوالات هر فرم نظر سنجی در یک فایل متنی مجزا
  + استخراج اطلاعات داده­ها و ساخت جداول مربوطه در پایگاه داده
* پیش پردازش داده­ها
  + حذف نویز
  + حذف داده­های پرت
  + تصمیم در مورد سوالات بی­پاسخ
* پیش پردازش متون
  + جداسازی کلمات
  + یکسان­سازی و نرمال­سازی
  + حذف ایست واژه­ها
* گزارشات آماری
  + ارائه انواع گزارش­های آماری از پاسخ نامه­ها
  + بررسی همبستگی میان پاسخ‌های سؤالات مختلف در یک پرسشنامه
  + ارائه انواع آزمون­های آماری
* داده کاوی
* ارائه الگوهای تکرار شونده پنهان در پاسخنامه­ها
* پیش بینی رفتار افراد بر اساس الگوهای استخراج شده
* یافتن افراد مشابه هم در جامعه و بیان مهمترین ویژگی­های آنها
* دسته­بندی داده­ها بر اساس ویژگی­های مشابه
* متن کاوی
* استخراج موضوعات اصلی نظرات متنی
* یافتن علایق ونظرات 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

## شرح علمی پروژه

برای انجام این پروژه سه رویکرد پژوهشی اصلی وجود دارد: استخراج اطلاعات آماری، داده‌کاوی و متن­کاوی.

با کمک روش‌های مرتبط با استخراج اطلاعات آماری و داده‌کاوی می‌توان به تحلیل پاسخ‌های ارائه‌شده به سؤالات چندگزینه‌ای پرداخت. درحالیکه تکنیک‌های مرتبط با متن‌کاوی برای تحلیل متونی به‌کار می‌روند که به‌عنوان نظرات و پیشنهادات در پاسخ سوالات آزاد توسط افراد مختلف ارسال شده‌اند.

### پیش­پردازش داده

برای پردازش داده­های محدودیت­های و چالش­های متفاوتی وجود دارد که پیش­پردازش را ضروری می­نماید. در اینجا سه مورد از مهمترین آنها ذکر می­گردد:

* یکی از آنها عدم پاسخ­دهی افراد به برخی از سوالات پرسشنامه است. به این حالت در علم داده­کاوی وجود داده از دست رفته می­گویند. در این شرایط یک راه حل نامناسب کنار گذاشتن آن پاسخنامه است که باعث می­شود سایر سوالاتی را هم که فرد جواب داده است کنار بگذاریم. بنابراین ما باید از روش­های جایگزین کردن جواب سوالات پاسخ داده نشده با جواب­های معتبر مناسب استفاده نماییم تا بتوانیم آن پرسشنامه را هم در تحلیل­ها در نظر بگیریم. اما یافتن مقدار مناسب برای پرسش­هایی که پاسخ داده نشده­اند نیازمند استفاده از الگوریتم­های متنوع و ارزیابی آنهاست که زمان بر خواهد بود.
* یکی دیگر از چالش­های پیش روی ما وجود داده­های پرت و اطلاعاتی است که ممکن است به اشتباه وارد شده باشد یا فردی عمدا اشتباه بگوید. وجود این داده­ها باعث انحراف الگوریتم­های داده کاوی خواهد شد و حتما می­بایست پیش از پردازش اصلی پرسشنامه­ها این داده­ها تشخیص داده و حذف شوند. که این مسئله خود باعث اضافه شدن مراحل پردازشی می­گردد.
* برقراری ارتباط بین پرسشنامه*­*­های مختلف نیز یکی از موارد نسبتا چالشی این پروژه است. ارتباط بین نظرات در پرسشنامه­هایی که در زمان­های مختلف در سازمان تدوین می­شود در صورتی می­تواند تحلیل گردند که برای موارد مشابه در زمان­های مختلف از نام یکسانی استفاده شده باشد. مثلا همیشه از لغت سن استفاده شود یا همیشه از تاریخ تولد و نه اینکه در یک پرسشنامه سن باشد و در دیگری تاریخ تولد. تنها در صورت وجود موارد مشابه همنام امکان تحلیل ترکیبی نتایج چند پرسشنامه وجود دارد.

### پیش­پردازش متن

همانند چالش­های مطرح در داده­کاوی چالش­های متن­کاوی نیز پیش­پردازش متون را ضروری می­نماید.

مهم‌ترین چالش مطرح در این پروژه مواجه‌شدن با متونی است که توسط افراد عامی جامعه تولید شده‌است. به‌عنوان مثال متون خبری و روزنامه‌ای و یا نگارش‌شده در کتب توسط گزارش‌گران و یا نویسندگانی تهیه می‌شود که از سطح دانش بالایی در زمینه اصول زبانشناسی و نگارشی برخوردارند. درحالیکه متون گردآوری‌شده در طرح‌های نظرسنجی و یا متون موجود در وبلاگ‌ها و شبکه‌های اجتماعی توسط افراد عامی جامعه نگارش می‌شود که دانش کافی برای رعایت استانداردهای نگارش را ندارند. از طرفی در این موارد متون رسمی و غیررسمی به‌صورت ترکیبی دیده می‌شود. این مسئله باعث بروز دو مشکل در پردازش متون می‌گردد: ۱) به‌دلیل عدم رعایت اصول نگارش صحیح، تشخیص مرز میان واژه‌ها که اصلی‌ترین واحد پردازش در متن کاوی است دشوار می‌گردد. چراکه تعداد زیادی از کلمات که باید متصل نوشته‌ شوند به‌صورت اشتباه جدا نوشته می‌شود. به‌عنوان مثال واژه «دست‌اندرکاران» ممکن است به اشتباه به‌صورت «دست اندر کاران» یا «دست اندرکاران» نوشته شود. همین‌طور تعداد زیادی از واژه‌ها به‌صورت اشتباه متصل به‌هم نوشته می‌شوند. مانند عبارت «ازتولیدواقتصادداخلی» که بدون هیچ فاصله‌ای بین واژه‌ها نگارش شده‌است. وجود این اشکال حتی می‌تواند ابهامات زیادی را وارد سیستم نماید. مانند واژه «مادر» که می‌تواند یک واژه مستقل باشد و یا از اتصال اشتباه دو واژه «ما» و «در» ایجاد شده باشد. ۲) به‌دلیل عدم رعایت استاندارد‌های نگارشی، تنوع نگارشی زیادی در متون و کلمات به‌کاررفته در آن‌ها مشاهده می‌گردد. به‌عنوان مثال ممکن است واژه «مسئولین» با نگارش‌های زیر ظاهر گردد:‌ «مسأولین»، «مسؤلین»، «مسولین»، «مسؤولین» و «مسئولین».

با توجه به مشکلات مطرح شده، باید در مرحله آغازین پروژه زیرساختی را برای پیش‌پردازش متون فراهم آورد تا بتوان درحدامکان اشتباهات نگارشی که باعث عدم تشخیص مرز صحیح واژه می‌شود را رفع نمود و همین‌طور  با تشخیص نگارش‌های مختلف واژه‌ها اقدام به یکسان‌سازی آنها کرد.

علاوه بر موارد فوق، در صورتی که داده­ها به­صورت دیجیتال جمع­آوری شوند با مشکلات دیگری نیز در نگارش متن فارسی مواجه خواهیم بود. یکی از این مشکلات، استفاده از حروف لاتین برای نگارش متن فارسی است. برای رفع این مشکل می‌بایست به تفکیک این متون در قسمت پیش‌پردازش پرداخت و درصورتی که این داده‌ها درصد قابل توجهی از متن را به‌خود اختصاص دهند اقدام به تبدیل خودکار آنها به خط فارسی و استفاده از آنها در بخش‌های تحلیلی نمود. مشکل دیگر وجود کدگزاری‌های مختلف برای متون فارسی می‌باشد که به‌دلیل تداخل حروف فارسی با زبان‌های عربی، دری و پشتو ایجاد شده‌است. وجود سیستم‌عامل‌های مختلف در رایانه و تلفن همراه به این مشکل دامن‌زده‌است. به‌گونه‌ای که در برخی از ساختارهای تعریف صفحه کلید حروفی مانند «گ» و «ژ» وجود ندارد و کاربر از حروف نزدیک به‌آنها مانند «ک» و «ز» استفاده می‌کند. این جابجایی سبب تولید لغات جدید و اشتباه نگارشی می‌گردد که باید در قسمت پیش‌پردازش مورد توجه و اصلاح قرار گیرد.

### استخراج اطلاعات آماری

استخراج آماره­های مختلف و ارائه آزمون­های آماری نظیر آزمون t-student، یکی از موارد مهم در این سیستم ارائه گزارش های آماری است. همچنیت هنگامی که آماره­ای در مورد یک موضوع گزارش می­شود مثلا بیان می­شود که در نتایج نظر سنجی بانوان به سریال­ها امتیاز مثلا 3.6 داده اند و آقایان 3.7 نمی توان به راحتی نتیجه گرفت مردان از سریال­ها راضی­تر هستند بلکه ممکن است درست­تر باشد که نتیجه بگیریم تفاوت معناداری بین نظر این دو گروه وجود ندارد به این منظور ما از آزمون­های آماری برای بررسی معناداری اختلاف نتایج استفاده خواهیم کرد. بنابراین در ماژول پردازش های آماری به استخراج اطلاعات آماری و بررسی آزمون های آماری جهت مشخص نمودن میزان اطمینان نتایج خواهیم پرداخت. این ماژول شامل موارد زیر خواهد شد:

1. امکان انتخاب و جستجوی یک یا چند پرسشنامه ها جهت انجام گزارشات آماری ( در صورتیکه ارتباط قبلا بین پرسشنامه ها تعریف شده باشد )
2. جدول های فراوانی تک تک پرسش ها که شامل فراوانی، درصد و درصد تراکمی باشد.
3. جدول های متقاطع بین پرسش ها ( حداکثر 5 مورد ) با امکان انتخاب فراوانی خانه ها، درصدکل، درصد سطری، درصد ستونی و درصدهای حاشیه ای.
4. امکان محاسبه مشخصه‌های توصیفی (میانگین، میانگین پیراسته، انحراف‌معیار، دامنه، مقدار کمینه، مقدار بیشینه، میانه، نما، چندک‌ها، چولگی و کشیدگی)
5. امکان محاسبه نمره کل از مجموعه‌ای از پرسش‌ها و تعیین دامنه نمره کل.
   1. نمره کل به صورت وزنی یا غیر وزنی باشد ( درصورت وزن نداشتن سوال، وزن یک به عنوان پیش فرض درنظر گرفته می شود.) کاربر باید بتواند برای سوالات وزن مورد نظر خود را تعریف نماید.
   2. شیوه محاسبه شاخص انتخابی باشد( میانگین یا جمع )
   3. بازه ی محاسبه شاخص انتخابی باشد.
6. امکان محاسبه ضرایب همبستگی پیرسون، اسپیرمن و تاو کندال و آزمون‌های آماری آنها.
7. امکان محاسبه ضرایب وی کرامر، دی سامرز و ضریب توافقی و آزمون‌های آماری آنها.
8. امکان محاسبه شاخص‌های اِتا دو و اُمگا دو.
9. امکان نمایش جدول و رسم نمودارهای دایره‌ای، میله‌ای، هسیتوگرام، پراکنش و خطی با امکان چاپ ( برای تمام موارد مندرج در این بخش ، همچنین پارامتر نمره کل نیز در تمام موارد به عنوان یکی از گزینه ها در محاسبات قرار داده شود.
10. امکان اجرای آزمون‌های آماری خی‌دو، تی، اِف (تجزیه واریانس)، کروسکال‌والیس، من‌ویتنی، میانه.
11. امکان اجرای آزمون گردش (تصادفی بودن نمونه، Run Test)
12. امکان محاسبه واریانس درون بلوکی و بین بلوکی و همبستگی درون خوشه‌ای (بلوکی)
13. امکان اجرای تمام محاسبات ذکرشده در زیرگروها ( مثلا محاسبه نمره کل برای گروه مرد ها و زن ها به طور مجزا )
14. با امکان تعریف پرسش تفکیک کننده جهت تفکیک گزارش های آماری
15. امکان ارائه تمام روش های سنجش همبستگی از قبیل:
    1. مدل معادلات ساختاری
    2. تحلیل عاملی ( تبینی و اکتشافی ) و آزمون های مرتبط شامل آزمون کرویت و KMO ( توسط برنامه factor loading را نمایش دهد )
    3. محاسبه ضریب پایایی با انتخاب مجموعه ای از سوالات ( همه انواع ) به انتخاب کاربر
    4. محاسبه ضریب همبستگی چند رشته ای و چند خانه ای با انتخاب مجموعه ای از سوالات به انتخاب کاربر
16. آزمون نرمالیتی ( انتخاب یک سوال و بررسی نرمال بودن مقادیر توزیع شده )
17. امکان Clustering براساس ویژگی های انتخابی ( به صورت اکتشافی یا غیراکتشافی ) بر اساس سوالات تعیین شده cluster انجام دهد.
18. امکان استفاده از روش های زیر برای جانهی (imputation) افراد دارای پرسش بی پاسخ به روش های:

* نزدیک ترین همسایه،
* جانشینی پرسش همبسته،
* جانشینی با انتخاب تصادفی،
* جانهی رگرسیونی، جانهی نسبتی، و
* جانهی میانگین به سه شکل ماینگین فرد، میانگین پرسش و میانگین پرسش تصحیح شده

### داده‌کاوی

تحلیل جامع نظر افراد نیازمند بررسی دقیق پاسخ همه سؤالات نظرسنجی و درنظرگرفتن ارتباطات بین همه پاسخ‌های یک فرد، ارتباط بین پاسخ­های افراد مختلف و همچنین پاسخ نامه­های مرتبط با هم است. این مسئله کمک خواهد کرد تا دید روشن و دقیق‌تری از نظرات افراد به‌دست آید. با استفاده از الگوریتم­های داده کاوی این امکان را خواهیم داشت تا به استخراج دانش مورد نیاز و کشف روابط پنهان میان موضوعات مختلف یک پرسشنامه بپردازیم. همچنین این امکان را خواهیم داشت تا افراد جامعه را براساس کلیت افکار  و نظرات‌شان دسته بندی نماییم و سپس در مورد این گروه­ها تحلیل های دقیقتر و پیش بینی هایی در مورد آینده انجام دهیم. علاوه‌بر این، همبستگی‌های بین پاسخ سؤالات مختلف می‌تواند ما را در تحلیل بهتر پاسخ‌ها کمک نماید. از موارد دیگری که می توان به آن اشاره نمود، این است که این امکان وجود دارد که کلمات کلیدی نظرات آزاد افراد به‌همراه پاسخ‌هایی که به سؤالات دیگر داده‌اند با یکدیگر ترکیب شود و بررسی کاملی بر روی پاسخ افراد صورت ‌پذیرد.

به منظور داده کاوی بر روی سوالات پرسشنامه­ها عملیات­های داده کاوری زیر صورت خواهد گرفت:

۱) 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 نسبتا بالای داده ها 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.

۲) 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 70% علاقمندی او به برنامه های تحلیلی و اخبار است.) این گام یک روش مناسب برای یافتن روابط مهم بین متغیرهای موجود در پایگاه‌داده‌های بزرگ است. برای این منظور از یکی از دقیق‌ترین الگوریتم‌های کاوش قوانین وابستگی استفاده خواهد شد. این الگوریتم از استراتژی [جستجوی اول-سطح](https://fa.wikipedia.org/wiki/%D8%A7%D9%84%DA%AF%D9%88%D8%B1%DB%8C%D8%AA%D9%85_%D8%AC%D8%B3%D8%AA%D8%AC%D9%88%DB%8C_%D8%A7%D9%88%D9%84_%D8%B3%D8%B7%D8%AD) 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.

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

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

5) علاوه بر موارد ذکر شده عملیات­های پیش پردازشی نیز پیش از شروع تحلیل­ها لازم است که در بخش 2-5-1 بیشتر توضیح داده شده­اند.

### متن‌کاوی

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 PLSI و LDA 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 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، پس‌از استخراج موضوعات اصلی و یافتن ارتباط آن‌ها با گروه‌های نظردهنده، به‌منظور تحلیل افکار مطرح حول هر موضوع ازلحاظ مثبت یا منفی‌بودن، 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.

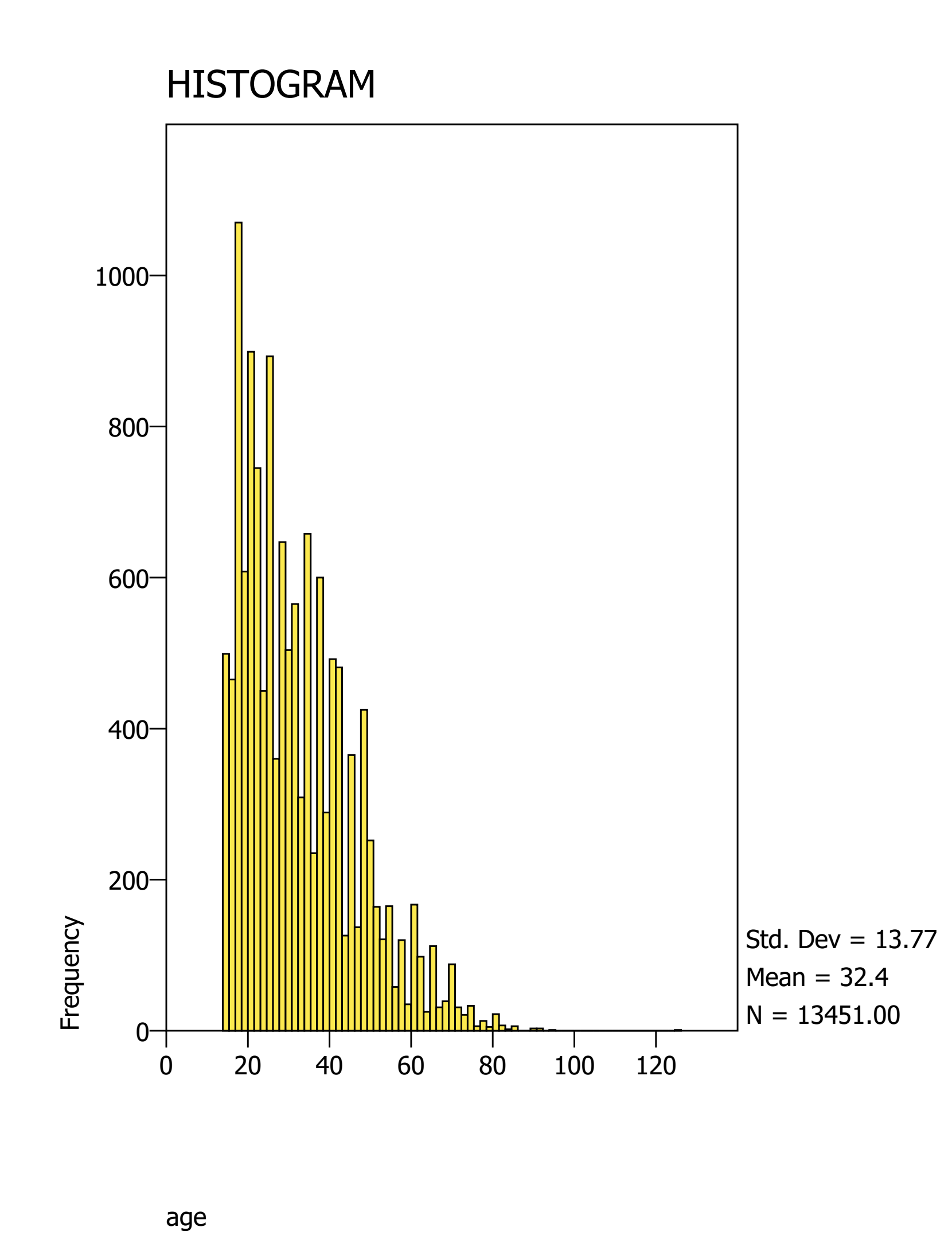
شایان ذکر است که بخش عظیمی از نظرات ارسال‌شده در این بخش جنبه پیشنهادی دارد و به‌طور مشخص حاوی نظر مثبت یا منفی حول موضوع مدنظر نمی‌باشد. تفکیک این نوع متون پیشنهادی از متونی که حاوی نظرات و عقاید خاص و مشخص هستند نیز در دستور کار این بخش از پروژه قراردارد.

## نمونه ساده از خروجی سیستم

مثال­های زیر چند نمونه ساده از خروجی سیستم براساس داده­های موجود برای نظرسنجی انتخابات می­باشد. این نمونه­ها با کمک روش­های اولیه استخراج شده­اند. بدیهی است با طراحی سیستم کامل و اتوماتیک روابط بیشتر و پیچیده­تری قابل استخراج خواهد بود.

* ارائه انواع گزارش­های آماری از پاسخ نامه­ها

به طور مثال ارائه هیستوگرام میانگین و انحراف معیار برای پاسخ همه سوالات. برای نمونه شکل زیر هیستوگرام متغیر سن را نشان می دهد.



* بررسی همبستگی میان پاسخ‌های سؤالات مختلف در یک پرسشنامه و یا در چند پرسشنامه مرتبط به هم

مثال بر اساس محاسبه همبستگی های دو به دو بین پاسخ های پرسش های مختلف چند نمونه از موارد به دست آمده در پرسشنامه در زیر ذکر شده اند:

* میان میزان شایستگی دکتر ولایتی، دکتر رضایی و دکتر حداد عادل همبستگی قوی وجود دارد. یعنی افرادی که مثلا دکتر ولایتی را شایسته می­دانند، دکتر حدادعادل را هم شایسته می­شمارند و همینطور افرادی که میزان شایستگی کمی برای دکتر ولایتی برای پست ریاست جمهوری قائل هستند، دکتر حدادعادل را نیز مناسب این پست نمی­دانند.
* همبستگی قابل توجهی میان نظر افراد در شایسته بودن یک فرد و احتمال رای آوردن او از دید آن فرد وجود دارد. یعنی اغلب افراد فردی را محتمل تر می­دانند که رای می­آورد که خود آن را شایسته­تر می­دانند.
* میزان رضایت از عملکرد صدا و سیما هیچ همبستگی معناداری با سن افراد ندارد.
* ارائه الگوهای تکرار شونده پنهان در پاسخ­نامه­ها (قوانین انجمنی)

چند نمونه از خروجی­هایی را که بر اساس تحلیل قوانین انجمنی به­دست آمده­است در زیر ارائه شده­اند.

* + تفاوت میزان مشارکت در انتخابات درمورد طرفدارن یک کاندیدا:به طور میانگین 60% افراد در انتخابات شرکت می­کنند. از میان طرفداران افراد مختلف نیز به طور میانگین حدود 60% در انتخابات شرکت می­کنند. یعنی به طور مثال 60% از افرادی که دکتر ولایتی را شایسته ریاست جمهوری می­دانند در انتخابات شرکت می­کنند. یا مثلا 60% از افرادی که دکتر عارف را شایسته ریاست جمهوری می­دانند در انتخابات شرکت می­کنند. اما این مسئله در مورد آیت­الله هاشمی متفاوت است و تنها 50% از طرفداران ایشان در انتخابات شرکت می­کنند. یعنی طرفداران ایشان 10% کمتر از میانگین طرفداران سایر کاندیداها در انتخابات شرکت می­کنند.
  + تفاوت رویکرد زنان تحصیلکرده و خانه­دار: زنان با تحصیلات لیسانس و بالاتر بیش از دو برابر نسبت به زنان خانه­دار به دکتر معین اقبال نشان می­دهند. و زنان خانه دار نزدیک بیش از 1.5 برابر زنان تحصیل کرده به دکتر لاریجانی اقبال نشان می­دهند.
  + توجه به بیکاری: مردان بیکار در مقایسه با سایر مردان اقبال بسیار کمتری به دکتر کروبی دارند. شاید این کاندیدا توجه کمتری به شعارهای اشتغال­زایی داشته است. این در حالی است که مردان شاغل تحصیل کرده بیش از سایر مردان به ایشان اقبال نشان داده­اند.
  + رضایت­مندی از صدا و سیما: میزان نارضایتی از عملکرد صدا و سیما در مردان بین 30 تا 40 سال 1.5 برابر مردان زیر بیست سال است.

در مجموع انتظار می­رود سیستم پیشنهادی بتواند قوانین و روابط مختلف مانند نمونه­های فوق و البته موارد پیچیده­تر از آن را با دقت بالای 90% پیش­بینی نماید. این سیستم همچنین قادر خواهد بود برای هریک از موارد هم­بستگی یک ضریب اطمینان نیز گزارش کند تا کاربر درک بهتری از میزان ارزش هریک از روابط داشته باشد. بدیهی است با داشتن تعداد بالاتری از نظرات در هر مجموعه داده می­توان این دقت را به­طور چشم­گیری افزایش داد.

## مشخصات فنی سیستم

در سیستم پیشنهادی موارد زیر مورد توجه قرار خواهد گرفت:

1. واسط کاربر مناسب
2. ارائه خروجی­های فارسی
3. سطوح دسترسی­های مختلف برای کاربران متفاوت
4. دسترسی همزمان چند کاربر در یک سامانه تحت وب
5. پشتیبانی سیستم

تخمینی از مشخصات سخت­افزاری سرور مورد نیاز پروژه به پیوست این پیشنهادنامه ارائه گردیده است.

## گسترش سیستم

در صورت توافق طرفین قرارداد سیستم پیشنهادی قابلیت گسترش 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)

## پشتیبانی سیستم

**الف\_ تعهدات پیمانکار :**

1. پیمانکار موظف است در یک بازه دوساله (پس از تحویل نرم افزار) به صورت رایگان و در بازه پنج ساله بعدی با دریافت هزینه (مطابق بند 7 بخش ب) نسبت به رفع هر گونه مشکل سیستم نرم افزار موضوع قرارداد اقدام نماید
2. پیمانکار متعهد مي­شود كه حسب درخواست مستند کارفرما ، خدمات پشتیبانی نرم افزاري مرتبط با نرم افزار كاربردي را به صورت راه دور با استفاده از امكانات مخابراتي، حداكثر ظرف مدت 24 ساعت به غير از ايام تعطيل به کارفرما ارائه نمايد .
3. در صورت  حل نشدن مشکل از راه دور پس از درخواست مستند كارفرما، پیمانکار موظف است حداكثر ظرف مدت 48 ساعت به غير از ايام تعطيل به محل استقرار نرم افزار كاربردي مراجعه و اقدام به رفع اشكال پيش آمده نمايد.
4. پیمانکار متعهد می­شود درخواست­های کارفرما برای توسعه سیستم را در مدت زمان پشتیبانی بررسی و برآورد هزینه نماید و درصورت توافق طرفین اقدام به توسعه نماید.

**ب\_ تعهدات کارفرما :**

1. کارفرما متعهد مي­شود شرايط سخت افزاري، نرم افزاري و شبکه مورد  نياز  را براي  نصب و اجراي  نرم افزار كاربردي  در طول  مدت قرارداد و پشتیبانی مهيا نمايد.
2. کارفرما متعهد است اشکالات احتمالی بوجود آمده را به­صورت مستند (از طریق ایمیل یا سیستم تیکتینگ) به اطلاع پیمانکار برساند.
3. در صورت درخواست کارفرما  براي حضور كارشناس پیمانکار در محل نصب نرم افزار كاربردي ، جهت انجام امور برنامه ريزي شده و غير قابل انجام از راه دور (با استفاده از تلفن یا ایمیل)، کارفرما مي بايست درخواست خود برای مراجعه حضوری پیمانکار را به­صورت مستند (از طریق ایمیل یا سیستم تیکتینگ) به پیمانکار اطلاع دهد.
4. کارفرما مي­بايست كتباً فردي را به عنوان رابط با پیمانکار خود در طول مدت قرارداد ،  معرفي و تا حد امكان از جابجايي وي اجتناب نمايد .
5. کارفرما متعهد مي­شود، بررسي صحت فرمت اطلاعات ورودی نرم افزار را همواره انجام دهد. در غیر اینصورت سیستم صرفا پیغام خطای مربوطه را نمایش خواهد داد و امکان پردازش اطلاعات را نخواهد داشت.
6. کارفرما مي­بايست كليه اطلاعات مورد نياز پیمانکار ( جهت ارائه خدمات پشتيباني ) را در اسرع وقت در اختيار وي قرار دهد. بدیهی است هرگونه تاخیر در ارائه اطلاعات منجر به تاخیر در پشتیبانی سیستم توسط پیمانکار خواهد گردید.
7. دربازه پنج ساله تامین هزینه هرگونه مشاوره و نگهداری سیستم اعم از تلفنی، اینترنتی و یا حضوری با نرخ زیر به عهده کارفرما می باشد.

در سال 1400 که شروع تعهد 5 ساله می­باشد:

* پشتیبانی از راه دور : هر ساعت 800000 ریال
* پشتیبانی حضوری : هر ساعت 1500000 ریال

به ازای سال­های بعد هر سال 10% به هزینه سال قبل افزوده خواهد شد.

## ساختار شكست پروژه

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| فاز | شماره فعاليت | شرح فعاليت | مدت اجرا  (روز) | مدت اجرا کل فاز  (روز) | درصد به کل پروژه | پیش‌نیاز |
| 1 | 1-1 | تهیه گزارش نیازسنجی پروژه  (مطابق با بخش 1 و 2  (به‌جز بندهای 2.1.3 و 2.1.4) | 15 | 15 | 30% | - |
| 2 | 2-۱ | سیستم استخراج اطلاعات داده­ها و ساخت جداول در پایگاه داده | 20 | 150 | 30% | - |
| 2-2 | سیستم یکسان­سازی فرمت داده­ها تبدیل داده­ها به فرمت لازم و ساخت پایگاه داده  (مطابق با بخش 1-2 مشخصات فنی پروژه در RFP )  (به جز مبحث Stored Prosedure ) | 15 |
| 2-3 | سیستم پیش­پردازش پاسخنامه­ها  (تشخیص نویز و حذف داده­های پرت و تصمیم­گیری در مورد سوالات بی­پاسخ) | 15 |
| 2-4 | سیستم ارائه انواع گزارش­های آماری از پاسخ نامه­ها  (مطابق با بندهای 1 تا20 بخش 1-4 مشخصات فنی پروژه در RFP) | 40 |
| 2-5 | واسط کاربر ساده بر مبنای پایگاه داده طراحی­شده و گزارشات آماری | 30 |
| 2-6 | سیستم مدیریت پرسشنامه­ها  (مطابق با بندهای 1 تا5 بخش 1-3 مشخصات فنی پروژه در RFP)  (به­جز بند 6 بخش 1-3 مشخصات فنی پروژه در RFP) | 5 |
| 2-7 | پیاده­سازی ماژول تولید خروجی­های عددی قابل استفاده در اکسل آفیس  (مطابق با بند 7 بخش 1-5 و بند 5 بخش 1-6 و بند 8 بخش 1-7 و بندهای 1 تا 3 بخش 1-8 مشخصات فنی پروژه در RFP)  (بجز ویژگی­های خط و زبان فارسی و خروجی تحت ورد آفیس  که در بند 6 بخش 1-5 مشخصات فنی پروژه در RFP ذکر شده) | 5 |
| 2-8 | طراحی و پیاده­سازی نسخه اولیه سامانه به همراه سطوح دسترسی  )مطابق با بندهای 1 تا 6 بخش 1-1 و بندهای 1 تا 7 بخش 1-7 مشخصات فنی پروژه در RFP) | 90 |
| 2-9 | تهیه گزارش فاز 2 | 5 |
| 3 | 3-1 | سیستم داده­کاوی  (مطابق با بندهای 1 تا 5 بخش 1-5 مشخصات فنی پروژه در RFP) | 60 | 120 | 25% | فاز 1 |
| 3-2 | سیستم پیش پردازش متون نظرات آزاد | 20 |
| 3-3 | سیستم متن کاوی  (مطابق با بندهای 1 تا 4 بخش 1-6 مشخصات فنی پروژه در RFP) | 60 |
| 3-4 | تکمیل واسط کاربر بر اساس فعالیت­های داده­کاوی و متن کاوی | 60 |
| 3-5 | سیستم ذخیره فایل­های پشتبان  )مطابق با بند 7 بخش 1-1 مشخصات فنی پروژه در RFP) | 5 |
| 3-6 | ارزیابی عملکرد سیستم | 30 |
| 3-7 | تهیه نسخه نهایی و یکپارچه ابزار خروجی  (به­جز بندا 8 بخش 1-1) | 30 |
| 3-8 | تهیه گزارش فاز 3  (مطابق با بخش 3 و 4  (به‌جز بندهای 3.2 و 4.1) | 5 |
| 4 | 4-1 | راه­اندازی سیستم در محل کارفرما | 10 | 45 | 15% | فاز 3 |
| 4-2 | رفع اشکالات عملی کار با سیستم | 10 |
| 4-3 | ارائه گزارش نهایی  (مطابق با بند 5.2 بخش 5 | 25 |