**Załącznik nr.1** Szczegółowy opis przedmiotu zamówienia.

**Opis przedmiotu zamówienia.**

**Przedmiotem postępowania jest dostawa licencji akademickiej na oprogramowanie do analizy statystycznej danych.**

Okres licencji: 31/03/2023 lub bezterminowe.

Licencja ma umożliwiać instalację oprogramowania w wirtualnym laboratorium statystyki (wirtualne stacje robocze pracujące na systemie Windows).

Szacowana ilość profili użytkowników w ramach wirtualnego laboratorium: 250 .

Funkcjonalności pakietu mają pozwalać na projektowanie i przygotowywanie rozbudowanych raportów analitycznych, gotowych do dystrybucji w sieci www z poziomu pakietu. Raporty mają być przygotowywane w formie dashboardów lub raportów szczegółowych, które można również zapisać do formatu .doc oraz .pdf.

Pakiet powinien posiadać dedykowane narzędzie umożliwiające automatyzację wykonywania zadań analitycznych oraz aktualizację i dystrybucję raportów analitycznych.

Pakiet powinien posiadać dedykowane rozwiązanie w postaci środowiska www, które służy do publikacji i udostępnianiu raportów analitycznych stworzonych w aplikacjach pakietu.

Pakiet powinien posiadać wbudowane narzędzia do prezentacji wizualnej danych na mapach Polski (z podziałem na województwa, powiaty, stolice województw). Pakiet ma zawierać mapy Polski w podziale na województwa i powiaty dedykowane dla tego oprogramowania.

Pakiet powinien posiadać samouczek obsługi oprogramowania w następujących obszarach: wczytywanie danych, operacje na danych, operacje na zmiennych, analiza i wykresy, edycja wyników, raportowanie, dystrybucja raportów, interfejs aplikacji.

Pakiet powinien posiadać wbudowane procedury niestandardowej wizualizacji, w tym: nakładany wykres słupkowy, wielowymiarowy wykres rozrzutu, wykres rozrzutu i rozkładu, Róża Nightingale, wykres radarowy, wykres kaskadowy, wykres wiolinowy, wykres hierarchiczny, mapa drzewa, dashboardy w tym: procent planu słupki, procent planu kostki, procent planu mierniki, procent planu progi termometry, procent planu progi słupki, wykres Marimekko, wykres bąbelkowy oraz możliwość formatowania obiektów dla wybranych procedur, zgodnie ze standardami APA.

Pakiet powinien posiadać wbudowane procedur do zarządzania danymi i wynikami, w tym: kodowanie dychotomiczne, słownikowy opis danych, akcje na raporcie, statystyki do stopek, inwentaryzacja danych, kolorowanie tabel, kopiowanie etykiet wartości, rekoduj kategorie monotoniczne, rekoduj kategorie mało liczne, przeskaluj zmienne.

Pakiet powinien posiadać wbudowane procedury dotyczące analizy danych w tym: istotne zmienne Chi-kwadrat, skorelowane zmienne V-Cramera, miary nierówności.

Pakiet powinien umożliwiać formatowanie wyników w raportach, w tym: dodatkowe szablony tabel, które formatują strukturę tabel przestawnych (wzbogacona paleta stylów).

Pakiet powinien umożliwiać formatowanie wyników w raportach, w tym: dodatkowe szablony tabel, które formatują strukturę tabel przestawnych (wzbogacona paleta stylów).

Pakiet powinien posiadać procedury w obszarze danych tekstowych: wyczyść tekst – pozwala na przekształcanie, czyszczenie i dostosowanie zmiennych tekstowych do podanych przez użytkownika wytycznych oraz procedura porównaj tekst – pozwala na wyliczenie miar opisujących podobieństwo ciągów tekstowych.

Funkcjonalności pakietu powinny pozwalać na projektowanie i przygotowywanie rozbudowanych raportów analitycznych, gotowych do dystrybucji w sieci www z poziomu pakietu. Raporty mogą być przygotowywane w formie dashboardów lub raportów szczegółowych, które można również zapisać do formatu .doc oraz .pdf. Do tworzenia raportów wykorzystywane są obiekty wynikowe tkj. tabele, wykresy, dashboardy. Użytkownik ma możliwość zarządzania oraz edycji obiektów przygotowanych w ramach pakietu, które są wykorzystywane w raporcie.

Pakiet powinien posiadać dedykowane narzędzie umożliwiające automatyzację wykonywania zadań analitycznych oraz aktualizację i dystrybucję raportów analitycznych. Zadania mogą być uruchamiane za pomocą harmonogramu zarówno cyklicznie jaki i jednorozowo.

Pakiet powinien posiadać dedykowane rozwiązanie w postaci środowiska www, które służy do publikacji i udostępnianiu raportów analitycznych stworzonych w aplikacjach pakietu. Raporty publikowane są w repozytorium bazodanowym portalu i stamtąd w postaci stron HTML są udostępniane uprawnionym osobom poprzez przeglądarkę internetową.

Pakiet powinien posiadać wbudowane narzędzia do prezentacji wizualnej danych na mapach Polski (z podziałem na województwa, powiaty, stolice województw).

Pakiet powinien umożliwiać udostępnianie kurs obsługi oprogramowania (samouczek) w tym: ponad 40 prezentacji multimedialnych w 8 obszarach: wczytywanie danych, operacje na danych, operacje na zmiennych, analiza i wykresy, edycja wyników, raportowanie, dystrybucja raportów, interfejs aplikacji.

Pakiet powinien posiadać wbudowane procedury niestandardowej wizualizacji, w tym: nakładany wykres słupkowy, wielowymiarowy wykres rozrzutu, wykres rozrzutu i rozkładu, Róża Nightingale, wykres radarowy, wykres kaskadowy, wykres wiolinowy, wykres hierarchiczny, mapa drzewa, dashboardy w tym: procent planu słupki, procent planu kostki, procent planu mierniki, procent planu progi termometry, procent planu progi słupki, wykres Marimekko, wykres bąbelkowy oraz możliwość formatowania obiektów dla wybranych procedur, zgodnie ze standardami APA.

Pakiet powinien posiadać wbudowane procedur do zarządzania danymi i wynikami, w tym: kodowanie dychotomiczne, słownikowy opis danych, akcje na raporcie, statystyki do stopek, inwentaryzacja danych, kolorowanie tabel, kopiowanie etykiet wartości, rekoduj kategorie monotoniczne, rekoduj kategorie mało liczne, przeskaluj zmienne.

Pakiet powinien posiadać wbudowane procedury dotyczące analizy danych w tym : istotne zmienne Chi-kwadrat, skorelowane zmienne V-Cramera, miary nierówności.

Pakiet powinien umożliwiać formatowanie wyników w raportach, w tym: dodatkowe szablony tabel, które formatują strukturę tabel przestawnych (wzbogacona paleta stylów).

Pakiet powinien umożliwiać formatowanie wyników w raportach, w tym: dodatkowe szablony tabel, które formatują strukturę tabel przestawnych (wzbogacona paleta stylów).

Pakiet powinien posiadać procedury w obszarze danych tekstowych: wyczyść tekst – pozwala na przekształcanie, czyszczenie i dostosowanie zmiennych tekstowych do podanych przez użytkownika wytycznych oraz procedura porównaj tekst – pozwala na wyliczenie miar opisujących podobieństwo ciągów tekstowych;

Oprogramowanie powinno zawierać następujące techniki analityczne: Analizę częstości, Statystyki opisowe, Eksplorację zmiennych, Tabele krzyżowe, Statystyki ilorazowe, Porównywanie średnich, Testy T Studenta, Jednoczynnikową ANOVA, Modele ANOVA - prosta analiza czynnikowa, Korelacje (parami, cząstkowe, odległości), Regresję liniową, Regresję porządkową, Estymację krzywej regresji, Testy nieparametryczne, Zestawy wielokrotnych odpowiedzi (tabele krzyżowe, częstości), Analizę czynnikową, Skalowanie wielowymiarowe (ALSCAL), Dwustopniowe grupowanie, Hierarchiczną analizę skupień, Analizę skupień metodą k-średnich, Analizę najbliższego sąsiedztwa, Analizę dyskryminacyjną, Kostki OLAP, Książkę kodów.

Oprogramowanie powinno pozwalać na szybką i efektywna kontrolę jakości danych. Przy pomocy jednego kreatora powinien pozwalać na zweryfikowanie zakresu danych w zbiorze, ilość braków danych w zmiennych czy powiązania logiczne pomiędzy zmiennymi. Oprogramowanie powinno korzystać z szeregu zdefiniowanych reguł lub konstruować własne. Oprogramowanie powinno automatycznie wychwytywać i raportować, tworzyć, filtrować obserwacje do analiz niespójności w danych.

Oprogramowanie powinno pozwalać również na szerokie wyszukiwanie wszelkiego typu anomalii w zbiorze danych, co może być wykorzystywane np. w kontroli i wykrywaniu nadużyć lub wykrywaniu nietypowych wzorów zachowań.

Oprogramowanie powinno umożliwiać badaczom i analitykom wykorzystać techniki bootstrappingowe. Technika bootstrappingu pozwala na oszacowanie rozkładu estymowanej wartości w toku procedury wielokrotnego losowania ze zwracaniem z oryginalnej próby, pozwalając na otrzymanie najbardziej wiarygodnego oszacowania błędów standardowych oraz przedziałów ufności takich parametrów populacji jak: średnia, mediana, proporcja, stosunek szans, współczynniki korelacji i regresji oraz wiele innych.

Oprogramowanie powinno posiadać możliwości prezentacji wyników za pomocą przejrzystych raportów tabelarycznych. Oprogramowanie powinno pozwalać na tworzenie dowolnych zestawień tabelarycznych. Oprogramowanie powinno pozwalać w elastyczny sposób na zarządzanie rodzajami zmiennych, przedstawianymi statystykami, robieniem podkategorii, sortowaniem wyników w tabeli włącznie z tworzeniem tabel dla zmiennych wielokrotnego wyboru.

Oprogramowanie powinno posiadać:

• prosty w obsłudze interfejs kreatora tabeli – obsługiwany przy pomocy mechanizmu „przeciągnij i upuść”. Zmienne przeciągane są do kolumn, wierszy czy warstw w zależności od potrzeb ich prezentacji,

• podgląd tworzonych tabel,

• możliwość tworzenia tabeli w trzech wymiarach: wiersze, kolumny i warstwy,

• możliwość krzyżowania w tej samej tabeli wielu niezależnych zmiennych,

• możliwość wyświetlania w jednej tabeli baterii pytań o takim samie zestawie odpowiedzi dla różnych pytań,

• możliwość umieszczania w tabeli zestawów wielokrotnych odpowiedzi,

• możliwość wyświetlania statystyk na raz w wierszach, kolumnach i warstwach (liczebności, % w wierszu, % w kolumnie, % w tabeli, średnia, mediana),

• możliwość jednoczesnego obliczenia następujących testów istotności: test niezależności chi-kwadrat, porównywanie proporcji w kolumnach, w kolumnach (test t), porównywanie wartości średnich,

• możliwość formatowania wszystkich elementów tabel.

Oprogramowanie powinno oferować użytkownikowi szereg technik wielowymiarowych takich jak: Ogólny model liniowy (GLM), Szacowanie składowych wariancji, MANOVA, Analiza Kaplana-Meiera, Regresja Coxa, Hierarchiczna analiza logliniowa, Analizy logliniowe i przeżycia.

Oprogramowanie powinno oferować zaawansowane modele regresji przydatne w wyjaśnianiu przyczyn występowania i przewidywaniu zjawisk, zdarzeń czy postaw, wykraczając poza zakres regresji liniowej i porządkowej, m.in.: regresja logistyczna – dwumianowa, regresja logistyczna, wielomianowa; logit i probit; regresja metodą ważonych najmniejszych kwadratów; regresja metodą cząstkowych najmniejszych kwadratów; regresja nieliniowa (przy warunkach pobocznych oraz bez war. pobocznych);

Oprogramowanie powinno umożliwiać wykorzystywanie technik do budowania map percepcyjnych, jednej z podstawowych technik obrazujących pozycję produkt na rynku. Zaimplementowane techniki powinny dawać możliwość poszukiwania "nisz" rynkowych, w których można by ulokować nasz nowy produkt lub przebudować produkt istniejący tak by zmienił swoją pozycję.

Oprogramowanie powinno dawać możliwość tworzenia czytelnych drzew klasyfikacyjnych  
i regresyjnych umożliwiających efektywną segmentację, grupowanie, predykcję, redukcję danych, identyfikację interakcji, łączenie kategorii oraz transformację zmiennych ciągłych w dyskretne. Powinno zawierać algorytmy do budowania drzew: CHAID, Exhaustive CHAID, C&RT, QUEST.

Oprogramowanie powinno pozwalać prowadzić wiarygodne badania na niewielkich próbach, pozwalać na dobór odpowiedniego testu statystycznego do posiadanych danych.

Oprogramowanie powinno pozwalać w prosty sposób otrzymać poprawne oszacowania błędu standardowego statystyk otrzymanych w próbie. Pozwalać losować próby o złożonych i wielostopniowych schematach doboru i zapisywać w odrębnym pliku wszystkie informacje o planie ich doboru.

Oprogramowanie powinno wspierać przygotowywanie prognoz poprzez analizę danych historycznych, budowanie modeli i prognozowanie przy wykorzystaniu pełnej mocy tkwiącej w module analizy szeregów czasowych.

Oprogramowanie powinno pozawalać na poszukiwanie powiązań pomiędzy brakami danych, a innymi zmiennymi w zbiorze danych, dodatkowo powinno dawać możliwość podstawiać za braki danych wartości oszacowane w oparciu o techniki regresyjną i EM.