

## Załącznik nr 1 do SWZ

### OPIS PRZEDMIOTU ZAMÓWIENIA

Oprogramowanie statystyczne dla pracowników i studentów Uniwersytetu Jana Kochanowskiego w Kielcach

Wykonawca obowiązany jest dostarczyć oprogramowanie wraz z licencją i kluczami do licencji.

Jeśli oprogramowanie będzie dostarczane na nośnikach wówczas w ramach gwarancji Zamawiający może domagać się od Wykonawcy nieodpłatnej wymiany nośników oprogramowania w terminie ustalonym przez obydwie strony.

Wykonawca obowiązany jest dostarczyć oprogramowanie i udostępnić do niego klucze w terminie do 10 dni roboczych od podpisania umowy.

#### 1. Cechy użytkowe oprogramowania:

- z oprogramowania mogą korzystać przez okres trzech lat do celów edukacyjnych i naukowo-badawczych pracownicy i studenci Uniwersytetu Jana Kochanowskiego w Kielcach (maks. 5000 stanowisk indywidualnych oraz instalacja sieciowa dla maks. 5000 jednocześnie użytkowników);
- główne środowisko pracy w programie w języku polskim (anglojęzyczne zwroty jedynie w niewielkim zakresie całego interfejsu);
- 36-miesięczna gwarancja, w ramach której Wykonawca zobowiązany jest do aktualizacji oprogramowania do nowych wersji oraz do świadczenia usługi pomocy technicznej dostępnej w dni robocze w godzinach od 8.00 do 16.00;
- Pomoc techniczna: to pomoc dla użytkownika, który ma problemy z oprogramowaniem (np. występują błędy, problemy z instalacją). Użytkownik w ramach pomocy może skontaktować się z pomocą techniczną drogą e-mailową lub telefonicznie, która pomoże rozwiązać problemy użytkownika.
- pomoc techniczna świadczona w języku polskim;
- rozbudowana pomoc elektroniczna zawierająca opisy poszczególnych opcji programu oraz dla wybranych modułów opisane krok po kroku przykłady analiz.

#### 2. Środowisko pracy z programem i korzystanie z zewnętrznych danych

- Dane mogą być składowane w arkuszu danych umożliwiającym interakcyjne wprowadzanie i przekształcanie danych (sortowanie, transformacje zmiennych, ułoż w stertę/rozrzuć po zmiennych) oraz import i eksport danych (m.in. z plików Excel i plików tekstowych).
- Oprogramowanie ma możliwość łączenia ze standardowymi bazami danych (SQL Server, MS Access i inne) przez OLE DB.
- Wczytywanie i zapis danych w formacie Excel (.xls, .xlsx, .xlsb, .xlsm), tekstowym, csv, html i innych.
- Wczytywanie i zapis plików danych w formatach: Statistica, SPSS, SAS, JMP, Minitab
- Oprogramowanie zawiera wbudowany, zgodny ze standardami język programowania Visual Basic, który umożliwia dostęp programowy do funkcji programu, programowanie własnych

procedur analitycznych (w tym węzłów analizy wykorzystywanych w przestrzeni roboczej data mining) oraz automatyzację prac.

- Środowisko użytkownika umożliwiające graficzne definiowanie projektu analitycznego w postaci schematu (grafu), w którym źródła danych, procedury przetwarzania danych i wyniki reprezentowane są przez ikony, a przepływ danych obrazują strzałki.
- Możliwość uruchamiania procedur w językach R i Python w projektach analitycznych zdefiniowanych jako schemat graficzny (graf).
- Oprogramowanie działa na stanowisku komputerowym pod kontrolą systemu operacyjnego Windows 10/11.

### 3. Zarządzanie wynikami

- Oprogramowanie zapewnia możliwość tworzenia raportów z analizy, z możliwością zapisania w formacie PDF.
- Przesyłanie wyników (tabel, wykresów) do dokumentów edytora tekstowego (np. MsWord).
- Możliwość ustawienia wynikowych tabel jako dane wejściowe dla kolejnych analiz.
- Raport otrzymywany przy pomocy Oprogramowania przypomina dokument edytora tekstu, a poszczególne obiekty (np. wykresy, arkusze, arkusz czy wykres MS Excel) umieszczane są w nim kolejno, jeden za drugim. Raporty mogą być zapisywane nie tylko we własnym formacie oprogramowania, ale także w postaci plików RTF, HTML.
- Oprogramowanie pozwala na zapis dokumentów (arkuszy danych i wyników, raporty) w postaci plików HTML, gotowych do opublikowania w Internecie lub Intranecie.
- Możliwość aktualizacji utworzonych wykresów po zmianie danych źródłowych automatycznie lub przez użytkownika (nie dotyczy map).
- Możliwość edycji wykresów po ich wstawieniu do dokumentu edytora tekstowego (tzn. wykresy mogą być wstawiane jako obiekty OLE) na komputerach z zainstalowanym Oprogramowaniem.

### 4. Funkcjonalność oprogramowania:

Oprogramowanie udostępnia w jednym środowisku użytkownika następujące funkcje analityczne:

- Statystyki podstawowe i tabele
- Możliwość wykonywania analiz w grupach
- Korelacje
- Kalkulator prawdopodobieństwa
- Dopasowanie rozkładów
- Regresja wieloraka
- Analiza wariancji (ANOVA)
- Statystyki nieparametryczne
- Wykresy: histogramy, wykresy rozrzutu, wykres workowy, wykresy średnia i błędy, wykresy ramka-wąsy, wykres składowych zmienności, wykresy zakresu, wykres rozrzutu z błędem, obrazkowe wykresy rozrzutu, wykresy rozrzutu z rysunkami, wykresy rozrzutu z histogramami, wykresy normalności, wykresy kwantyl-kwantyl, wykresy prawdopodobieństwo-prawdopodobieństwo, wykresy słupkowe/kolumnowe, wykresy liniowe, wykresy sekwencyjne/nakładane, wykresy kołowe, wykresy brakujących danych i spoza zakresu, histogramy dwóch zmiennych, wykresy powierzchniowe, wykresy warstwiczne, wykresy wafłowe, wykresy trójkątne, skategoryzowane wykresy XYZ,

skategoryzowane wykresy trójkątne, wykresy macierzowe, wykresy obrazkowe, wykresy XYZ 3W, wykresy trójkątne 3W

- Rozkłady i symulacje
- Ogólne modele liniowe
- Uogólnione modele liniowe i nieliniowe
- Ogólne modele regresji
- Modele częściowych najmniejszych kwadratów
- Komponenty wariancyjne
- Analiza przeżycia
- Estymacja nieliniowa
- Linearyzowana regresja nieliniowa
- Analiza log-liniowa tabel licznosci
- Szeregi czasowe i prognozowanie
- Modelowanie równań strukturalnych
- Analiza skupień
- Analiza czynnikowa
- Składowe główne i klasyfikacja
- Algorytm NIPALS dla analizy składowych głównych i metody częściowych najmniejszych kwadratów
- Analiza kanoniczna
- Analiza rzetelności i pozycji
- Drzewa klasyfikacyjne
- Analiza korespondencji
- Skalowanie wielowymiarowe
- Analiza dyskryminacyjna
- Ogólne modele analizy dyskryminacyjnej
- Analiza Mocy Testów
- Sieci neuronowe
- Dobór i eliminacja zmiennych (dla dużych zbiorów danych)
- Analiza koszykowa
- Interakcyjne drążenie danych
- Analiza skupień uogólnioną metodą EM i k-średnich
- Uogólnione modele addytywne (GAM)
- Ogólne modele drzew klasyfikacyjnych i regresyjnych (GTrees)
- Ogólne modele CHAID (Chi-square Automatic Interaction Detection)
- Interakcyjne drzewa klasyfikacyjna i regresyjne
- Wzmacniane drzewa klasyfikacyjne i regresyjne (Boosted Rrees)
- Multivariate Adaptive Regression Splines (MAR Splines)
- Obliczanie dobroci dopasowania
- Szybkie wdrażanie modeli predykcyjnych
- Naiwny klasyfikator Bayesa
- Support Vector Machines
- Metoda k-najbliższych sąsiadów
- Łączenie grup (klas) z wykorzystaniem algorytmu CHAID
- ICA (Independent Component Analysis)

- Losowy las (Random Forests)
- Standardowe karty kontrolne: karta  $\bar{X}$  średniego i R, karta  $\bar{X}$  średniego i S, karta pojedynczych obserwacji i ruchomego rozstępu (I/MR), karta sum skumulowanych (CUSUM), karta średniej ruchomej (MA), karta wykładniczo ważonej średniej ruchomej (EWMA), karty dla pomiarów alternatywnych (C, U, Np, P), karta Pareto, karty wielowymiarowe, karty wielotorowe
- Interaktywne zaznaczanie i etykietowanie punktów
- Przypisywanie przyczyn i działań
- Elastyczny, dostosowywalny system alarmowania
- Praca inżyniera i operatora; zabezpieczanie hasłem
- Karty krótkich serii
- Karty wieloźródłowe (zgrupowane i zgrupowane krótkich serii)
- Wskaźniki zdolności, wykonania i linie kontrolne dla rozkładów innych niż normalny
- Karty kontrolne w czasie rzeczywistym; zewnętrzne źródła danych
- Wielowymiarowe karty kontrolne Kart  $T^2$  Hotellinga
- Wielowymiarowe karty kontrolne Wieloźródłowych (zgrupowanych) kart  $T^2$  Hotellinga
- Wielowymiarowe karty kontrolne wykładniczo ważonej średniej ruchomej (MEWMA)
- Wielowymiarowe karty sum skumulowanych (MCUSUM)
- Karta uogólnionej wariancji
- Analiza zdolności procesu: wskaźniki zdolności procesów (np. Cp, Cr, Cpk, Cpl, Cpu, K, Cpm, Pp, Pr, Ppk, Ppl, Ppu i inne),
- Plany badania i analiza powtarzalności i odtwarzalności pomiarów (R&R)
- Analiza Weibulla
- Analiza doświadczenia: Ogólne możliwości
- Analiza resztowa i przekształcenia
- Optymalizacja pojedynczej lub wielu wielkości wyjściowych:
- Standardowe plany frakcyjne dwuwartościowe  $2(k-p)$
- Plany frakcyjne  $2(k-p)$  o najmniejszej aberracji i maksymalnym niewywikłaniu
- Plany eliminacyjne (Placketta-Burmana)
- Plany frakcyjne trójwartościowe typu  $3(k-p)$  z podziałem na bloki oraz plany Boxa-Behnkena
- Plany centralne kompozycyjne (powierzchnia odpowiedzi)
- Plany kwadratów łacińskich
- Doświadczenia wg metody Taguchi
- Plany dla mieszanin i powierzchni o podstawie trójkątnej
- Plany dla ograniczonych powierzchni i mieszanin
- Plany D i A-optymalne
- Funkcjonalność text mining
- Analiza dokumentów zapisanych w formacie MS Word
- Zliczanie wystąpień słów
- Różne miary częstości występowania słów : prosta częstość, częstość binarna (ang. binary frequency), odwrotna częstość dokumentowa (ang. inverse document frequency), częstość logarytmiczna
- Możliwość określania własnej stop-listy
- Możliwość określania synonimów

- Wykonywanie rozkładu według wartości osobliwych (ang. singular value decomposition) dla miar częstości występowania słów w zbiorze dokumentów
- Analiza podstawowych przyczyn
- Optymalizacja wielkości wyjściowych
- Ogólna optymalizacja
- Wdrażanie modelu MSPC
- Analiza składowych głównych (PCA)
- Częstkowe najmniejsze kwadraty (PLS)
- Wielokierunkowe cząstkowe najmniejsze kwadraty wg partii (BMPLS)
- Wielokierunkowa analiza składowych głównych według czasu (TMPCA)
- Wielokierunkowe cząstkowe najmniejsze kwadraty wg czasu (TMPLS)
- Wykrywanie reguł asocjacji
- Analiza sekwencji
- Analiza skojarzeń
- Wykresy zmienności,
- Wykresy wielokrotne, pozwalające bezpośrednio porównywać wiele zmiennych zależnych,
- Komponenty wariancyjne z przedziałami ufności,
- Elastyczne operowanie wieloma zmiennymi zależnymi: jednoczesne analizowanie wielu zmiennych wg tego samego lub różnych planów,
- Wykresy komponentów wariancyjnych
- Tabele raportujące
- ETL
- Konstruktor reguł
- WoE (Weight of Evidence – waga dowodu)
- Reguły poprawności danych
- Analiza brakujących danych
- Przekodowanie na zmienne sztuczne
- Szybkie rekodowanie
- Przekształcenia zmiennych
- Zliczanie wystąpień
- Porządkowanie zmienne wielokrotnych odpowiedzi
- Kalkulator liczebności próby
- Ważenie wieńcowe przypadków
- Propensity score matching
- Podział na podpróby
- Podsumowanie skali pozycyjnej
- Podsumowanie skali rangowej
- Wykres dyferencjału semantycznego
- Wykres dla skali Stapela
- Rzetelność skali
- Metoda ocen porównawczych Thurstone'a
- Współczynniki zgodności sędziów
- Testy dla pojedynczej zmiennej
- Badanie istotności różnic
- Krzywe ROC

- Metaanaliza i metaregresja
- Kreator regresji logistycznej
- Kreator regresji liniowej
- Analiza conjoint
- Analiza aglomeracji
- Analiza PROFIT
- Uogólniona metoda składowych głównych (PCA)
- Porządkowanie liniowe
- Bootstrap
- Układy naprzemienne badań klinicznych
- Układ naprzemienny prosty – test parametryczny
- Układ naprzemienny prosty – test nieparametryczny
- Układ naprzemienny podwójny – test parametryczny
- Układy czynnikowe (split-plot, split-block i bloki randomizowane)
- Wykres Blanda-Altmana
- Regresja Passinga-Babloka i Deminga
- Wykres górkowy
- Wykres Youdena
- Granice wykrywalności
- Miary powiązania/efektów dla tabel 2x2
- Analiza koncentracji
- Standaryzowane miary efektu
- Test post hoc ANOVA Friedmana
- CATANOVA
- Karta CUSUM ważona ryzykiem
- Indeks KMO oraz Test sferyczności Bartletta
- Konfiguracyjna analiza częstości (CFA)
- Badanie ciągów pomiarów
- Przedziały odniesienia
- Przedział ufności dla ilorazów met. Fiellera
- Profile ryzyka
- Wielowymiarowe testy normalności
- Prosta regresja ortogonalna
- Badanie rozkładu empirycznego
- Wykres słupkowy (kolorowe słupki)
- Wykres sekwencyjny
- Wykres radarowy
- Wykres mozaikowy
- Wykres kołowy (SPie plot)
- Piramida populacyjna
- Diagram sieci neuronowej
- Wykres sieci Kohonena
- Wykres diamentowy
- Szereg czasowy z limitami
- Róża wiatrów

- Wykres Likerta
  - Wykres wrażliwości
  - Wykres słonecznikowy
  - Wykres skrzypcowy
  - Wykres piramidowy
  - Wykres motylkowy
  - Zapisz do pliku Excel
  - Zapisz pliki graficzne
  - Formatuj arkusz
  - Możliwość wizualizacji danych na mapach:
- a) Gotowe szablony map dostępne w programie obejmują podział Polski na: województwa, powiaty, gminy, okręgi wyborcze, województwa w podziale na powiaty, województwa w podziale na gminy, województwa w starym podziale
  - b) Możliwość wczytywania innych niż zawarte w programie szablonów map w formacie \*.shp
  - c) Kolorowanie na mapach tła obszarów wartościami zadanej zmiennej (predefiniowane palety do wyboru, możliwość ustalenia palety użytkownika, możliwość ustalenia własnych granic dla przedziałów legendy, możliwość zapisu/wczytania palety kolorów z/do pliku)
  - d) Generowanie wykresów kołowych i słupkowych (możliwość ręcznej zmiany wielkości wykresu, możliwość ręcznego ustalenia jego położenia, możliwość zmiany skalowania wysokości słupka względem wiersza/kolumny/całości, zmienny promień wykresu kołowego zależny od wartości ze zmiennej)
  - e) Wyświetlanie etykiet tekstowych pobranych z zadanej zmiennej lub zmiennej zawierającej mapowanie elementów wraz z formatowaniem zadanych przez użytkownika (kolor, krój itp.), oraz ręczną korektą położenia etykiety względem innych elementów wykresu
  - f) Różne stany wyświetlania elementów obszaru – aktywny, nieaktywny, ukryty
  - g) Rodzaj i grubość linii rysowanych jako granice może być zmieniana przez użytkownika
  - h) Możliwość zapisu/odczytu z i do pliku wszystkich opcji wyglądu mapy
  - i) Możliwość ręcznej edycji przez użytkownika szablonów map wczytanych w programie (usuwanie obszarów, scalanie obszarów) i zapisu jako nowy szablon
  - j) Możliwość zarejestrowania wygenerowanej mapy (z wizualizacją danych) w postaci makra
    - Dedykowane narzędzia do budowy modeli scoringowych za pomocą regresji logistycznej
    - Możliwość budowy modelu logistycznego na podstawie prób bootstrapowych
    - Budowa modelu typu SURVIVAL
    - Analiza wniosków odrzuconych:
      - Parceling
      - metoda k-najbliższych sąsiadów
    - Ranking predyktorów na podstawie miar Information Value, Gini oraz V Cramera
    - Narzędzie do wykrywania reguł i interakcji za pomocą metody losowy las
    - Generowanie rankingu interakcji pomiędzy parami zmiennych przy użyciu regresji logistycznej
    - Narzędzia do dyskretyzacji zmiennych na potrzeby modeli scoringowych – manualne i automatyczne definiowanie przedziałów dla zmiennych ciągłych oraz rekategoryzacja zmiennych jakościowych
    - Diagnozowanie jakości podziału na przedziały na podstawie WoE (weight of evidence), wskaźnika Information Value oraz odpowiednich wykresów
    - Możliwość uwzględnienia braków danych jako wartości nietypowych

- Manualne definiowanie przedziałów dla zmiennej ciągłej
- Manualne grupowanie dla zmiennej dyskretnej
- Automatyczne tworzenie przedziałów dla zmiennej ciągłej według zadanych parametrów dotyczących liczebności przypadków w poszczególnych przedziałach
- Automatyczne tworzenie przedziałów dla zmiennej dyskretnej na podstawie minimalnej liczności
- Automatyczne tworzenie przedziałów dla zmiennej ciągłej lub dyskretnej za pomocą algorytmu CHAID
- Obsługa wartości nietypowych
- Wybór reprezentantów skupisk skorelowanych zmiennych ilościowych za pomocą analizy głównych składowych
- Zapis definicji kategoryzacji zmiennych w plikach XML
- Możliwość wczytania skryptu dyskretyzacji XML i reedycja zdefiniowanych przedziałów
- Ocena jakości zbudowanych modeli na podstawie miar: Information Value, Kolmogorowa-Smirnowa, Hosmera-Lemeshowa, Dywergencji, Giniego, pola pod krzywą ROC
- Narzędzia do optymalizacji punktu odcięcia dla modeli scoringowych
  - możliwość wyboru punktu odcięcia (cut-off) na podstawie analizy ROC oraz kosztów błędnych klasyfikacji
  - możliwość wskazania od 1 do 3 punktów odcięcia
  - zestaw narzędzi i raportów pozwalających ocenić trafność odcięcia
  - możliwość wyboru punktu odcięcia na podstawie jednostkowych bądź średnich kosztów (np. wyrażonych kwotowo)
- Możliwość symulacji zyskowności modelu scoringowego dla wczytanego portfela w zależności od przeznaczenia modelu i podanych przez użytkownika parametrów dodatkowych
- Narzędzia do badania stabilności populacji i cech
- Zapis zbudowanego modelu scoringowego w postaci tablicy/karty scoringowej
- Zapis tablicy scoringowej w postaci arkusza Excel
- Generowanie raportu opisującego powstałą kartę scoringową
- Raporty: cech (characteristic report), końcowej punktacji (final score report), wykresy Bad rate oraz Odds
- Obliczenie wartości scoringu dla nowych danych na podstawie zbudowanych modeli scoringowych
  - obliczanie scoringu dla nowych danych na podstawie wybranego modelu
  - możliwość obliczania PD (default probability)
  - skalowanie wartości PD dla modeli budowanych na zbalansowanym zbiorze danych
  - obliczanie prawdopodobieństwa dla modeli typu SURVIVAL
- Ocena kart scoringowych zapisanych w postaci XML
- Ocena modeli na podstawie scoringu bądź prawdopodobieństwa zapisanego w arkuszu danych
- Analiza lift: wykres lift, wykres gain, raport wartości lift