**Opis Przedmiotu Zamówienia**

**Dostawa licencji zaawansowanego oprogramowania do analizy danych**

**Zaoferowania licencja dotyczy oprogramowania:………………………**

**Wersje:……………………………………………………………………………………**

Przedmiotem zamówienia jest dostawie klucza licencyjnego dla zaawansowanego oprogramowania do analizy danych oraz pliku licencyjnego dla poszczególnych modułów tego oprogramowania, kompatybilnego z oprogramowaniem posiadanym przez Zamawiającego**.** Wykonawca jest zobowiązany dostarczyć licencję w wersji jednostanowiskowej oraz sieciowej.

Zamawiający posiada następujące oprogramowanie wraz z modułami: Statistica w wersjach 12.5, 13.0, 13.1, 13.3 zainstalowanej obecnie na komputerach użytkowników Uniwersytetu Medycznego w Łodzi oraz jej modułów: Rozszerzony Pakiet Akademicki, Zestaw PLUS, Zestaw Skoringowy, Zestaw Mapy, Zestaw Przyrodnika, Zestaw Farmaceutyczny(obejmujący Profile Uwalniania, Ocena Stabilności, Walidacja Procesu, Walidacja Metod Pomiarowych, Przegląd Jakości Produktu), Zestaw Kliniczny.

Licencje będą obowiązywały przez **24 miesiące począwszy od 1 lutego 2022 r.**

Zaawansowane oprogramowanie do analizy danych musi umożliwiać wizualizowanie danych, zarządzanie nimi oraz eksplorowanie. Musi umożliwiać:

1. Wykorzystanie podstawowych technik analitycznych:

* Statystki podstawowe i tabele
* Wykonywanie analiz w grupach
* Korelacje
* Kalkulator prawdopodobieństwa
* Dopasowanie rozkładów
* Regresja wieloraka
* Analiza wariacji
* Statystki nieparametryczne
* Komplet narzędzi graficznych – wykresy:

histogra-my, 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 warstwicowe, wykresy waflowe, wykresy trójkątne, skategoryzowane wykresy XYZ, skategoryzowane wykresy trójkątne, wykresy macierzowe, wykresy obrazkowe, wykresy XYZ 3W, wykresy trójkątne 3W.

1. Wykorzystanie zaawansowanych metod statystycznych takich jak:

* Rozkłady i symulacje
* Ogólne modele liniowe
* Uogólnione modele liniowe i nieliniowe
* Ogólne modele regresji
* Modele cząstkowych najmniejszych kwadratów
* Komponenty wariancyjne
* Analiza przeżycia
* Estymacja nieliniowa
* Linearyzowana regresja nieliniowa
* Analiza logliniowa tabel liczności
* Szeregi czasowe i prognozowanie
* Modelowanie równań strukturalnych
* Analiza skupień
* Analiza czynnikowa
* Składowe główne i klasyfikacja
* Algorytm NIPALS dla analizy składowych głównych i metody cząstkowych najmniejszych kwadratów
* Analiza kanoniczna
* Analiza rzetelności i pozycji
* Drzewa klasyfikacyjne
* Analiza korespondencji
* Skalowanie wielowymiarowe
* Analiza dyskryminacyjna
* Ogólne modele analizy dyskryminacyjnej
* Analiza Mocy Testów

1. Korzystanie z funkcjonalności oprogramowania takich jak:

* 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 (splitplot, splitblock i bloki randomizowane)
* Wykres BlandaAltmana
* Regresja PassingaBabloka 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

1. Korzystanie z dedykowanych narzędzi wspierających proces budowy modeli skoringowych, w szczególności w postaci karty skoringowej, takich jak:

* 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:

o parceling

o 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 skoringowych – 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, Kołmogorowa Smirnowa, HosmeraLemeshowa, Dywergencji, Giniego, pola pod krzywą ROC
* Narzędzia do optymalizacji punktu odcięcia dla modeli scoringowych

o możliwość wyboru punktu odcięcia (*cut-off*) na podstawie analizy ROC oraz kosztów błędnych klasyfikacji

o możliwość wskazania od 1 do 3 punktów odcięcia

o zestaw narzędzi i raportów pozwalających ocenić trafność odcięcia

o 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

o obliczanie skoringu dla nowych danych na podstawie wybranego modelu

o możliwość obliczania PD (*default probability*)

o skalowanie wartości PD dla modeli budowanych na zbalansowanym zbiorze danych

o obliczanie prawdopodobieństwa dla modeli typu SURVIVAL

* Ocena kart scoringowych zapisanych w postaci XML
* Ocena modeli na podstawie scoringu bądź prawdopodobieństwa zapisanego w arkuszu danych
* Ocena jakości zbudowanych modeli na podstawie miar: IV (Information Value), KS (Kołmogorowa Smirnowa), HosmeraLemeshowa, Dywergencji, Giniego, Pola pod krzywą ROC
* Analiza lift: wykres lift, wykres gain, raport wartości lift

Oprogramowanie musi dać możliwość wizualizacji danych na mapach:

1. Gotowe szablony map obejmujące podział Polski na: województwa, powiaty, gminy, okręgi wyborcze, województwa w podziale na powiaty, województwa w podziale na gminy, województwa w starym podziale
2. Możliwość wczytywania innych niż zawarte w programie szablonów map w formacie \*.shp
3. 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)
4. 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)
5. Wyświetlanie etykiet tekstowych pobranych z zadanej zmiennej lub zmiennej zawierającej mapowanie elementów wraz z formatowaniem zadanych przez użytkownika (kolor, krój itp.), oraz ręczną korekcją położenia etykiety względem innych elementów wykresu
6. Różne stany wyświetlania elementów obszaru – aktywny, nieaktywny, ukryty
7. Rodzaj i grubość linii rysowanych jako granice może być zmieniana przez użytkownika
8. Możliwość zapisu/odczytu z i do pliku wszystkich opcji wyglądu mapy
9. Możliwość ręcznej edycji przez użytkownika szablonów map wczytanych w programie (usuwanie obszarów, scalanie obszarów) i zapisu jako nowy szablon
10. Możliwość zarejestrowania wygenerowanej mapy (z wizualizacją danych) w postaci makra

Oprogramowanie musi zawierać następujące moduły:

Moduł obejmujący:

* Statystyki podstawowe
* Układy ANOVA (jedno-, dwu- oraz trójczynnikowe)
* Kreator regresji liniowej
* Analiza aglomeracji
* AMOVA
* Doświadczenia hodowlane (linia x tester, Diallel kwadratowy, Diallel trójkątny, Diallel kwadratowy bez przekątnej, Diallel trójkątny bez przekątnej)
* Doświadczenie jednopowtórzeniowe z wzorcem
* Serie doświadczeń (split-split-plot, układ hierarchiczno-krzyżowy, AMMI),
* Doświadczenia kratowe
* Testy dla pojedynczej zmiennej
* Badanie istotności różnic

Moduł wykorzystywany jako podstawowe narzędzie w kontroli jakości a w szczególności:

* określenia różnic w zmienności wewnątrz i pomiędzy seriami
* rozróżnienia pomiędzy „złą” a „dobrą” serią produktu leczniczego
* przewidywania zachowania się produktu in-vivo
* oceny „istotnego podobieństwa” między produktem badanym a referencyjnym
* oceny „istotnego podobieństwa” pomiędzy produktami przed i po wprowadzeniu zmian w trakcie cyklu życia

Oprogramowanie ma umożliwiać dobór odpowiedniej metody porównania profilu dla danego przypadku, standaryzację przeprowadzanych analiz statystycznych i interpretację wyników oraz pozwala na wygenerowanie raportu w formie tabelarycznej i/lub graficznej koniecznego do złożenia dokumentacji rejestracyjnej. Metody wykorzystane w module mają być zgodne z rekomendacjami zawartymi w regulacjach agencji europejskiej EMA oraz amerykańskiej FDA:

* Metody porównania profili niezależne od modelu
* Współczynnik podobieństwa f1
* Współczynnik różnicy f2
* Wieloczynnikowa odległość statystyczna (oparta na odległości Mahalanobisa)

Metody porównania profili zależne od modelu:

* Model zerowego rzędu
* Model pierwszego rzędu
* Model Hixson – Crowella
* Model Higuchiego
* Model kwadratowy
* Model Gompertza
* Model logistyczny
* Model Weibulla
* Model Korsemeyara- Peppasa

Oprogramowanie musi dać możliwość tworzenie standardowych raportów do dokumentacji:

* Informacje o badaniu
* Zestawienie danych surowych
* Zestawienie statystyk opisowych
* Zestawienie parametrów wybranego modelu
* Ocena dobroci dopasowania modelu
* Graficzna analiza porównawcza
* Ocena podobieństwa profili uwalniania

Moduł musi umożliwiaćwyznaczanie okresu ważności dla pojedynczej serii jak i wielu serii dla jednego czynnika zgodnie z rekomendacjami ICHQ1E.B1 i ICHQ1E.B2 oraz do trendowania wyników gromadzonych podczas ciągłego badania stabilności wraz z oceną prawdopodobieństwa wystąpienia wyniku OOS na koniec okresu ważności OOS. Możliwość generowania okresowych dwóch rodzajów raportów. Pierwszy z nich dotyczy analizy okresu ważności produktu leczniczego, natomiast drugi pozwala na sprawdzenie trendu w analizowanej serii na podstawie jej wcześniejszych wyników lub z wykorzystaniem serii historycznych.

Podstawowe funkcjonalności systemowe mają obejmować:

* Przeprowadzenie analizy wielu serii dla wielu parametrów
* Konfigurację zawartości raportu z wyborem analiz statystycznych
* Zdefiniowanie standardowych specyfikacji badania
* Tworzenie typowych szablonów raportów zwierających pełną konfigurację analiz
* Generowanie raportu w formacie PDF/Word/inne po wprowadzeniu danych
* Definiowanie poziomów istotności dla analiz statystycznych
* Definiowanie kryteriów identyfikacji wyników OOT i OOS

Dostęp do wyników wielu metod statystycznych pod jednym przyciskiem np.:

* statystyki opisowe,
* karty kontrolne,
* ocena normalności danych,
* wskaźniki zdolności,
* ANOVA.

Możliwość wykonywania analizy statystycznych dla kilku serii.

Możliwość wprowadzenia danych do arkusza i naciśnięciu jednego przycisku w celu otrzymania raportu w postaci dokumentu Microsoft Word.

Oprogramowanie musi dać możliwość generowania sześciu rodzajów raportów:

* Sprawozdanie z eksperymentu walidacyjnego wg normy PN-ISO 5725
* Porównanie różnicy wartości średnich w dwóch zbiorach (np. porównanie metod)
* Wyznaczanie limitu detekcji i oznaczalności (LOD, LOQ)
* Kalibracja (liniowość)
* Regresja (odzysk)
* Porównanie prostych regresji

Automatyczne tworzenie raportu podsumowującego zdolność jakościową parametrów produktu lub parametrów środowiskowych. Zadaniem programu ma być podsumowanie wskaźników zdolności dla analizowanych parametrów i ocena istnienia niepożądanych trendów lub autokorelacji dla parametrów produktu lub parametrów środowiskowych. Program ma zapewnić funkcje potrzebne podczas tworzenia APR (*Annual Product Review*) i analizy trendów dla parametrów środowiskowych i został tak zaprojektowany aby wspomagać analizę danych pod kątem wytycznych zawartych w dokumencie: „*Rozporządzenie Ministra Zdrowia z dnia 23 grudnia 2010 r. zmieniające rozporządzenie w sprawie wymagań Dobrej Praktyki Wytwarzania*”.

W raporcie końcowym mogą znaleźć się następujące elementy:

* statystyki opisowe podsumowujące analizowane dane
* ocena wartości odstających od trendu w oparciu o odległość Cooka
* ocena istnienia trendu w oparciu o test Neumanna
* ocena istnienia trendu w oparciu o karty kontrolne Shewharta
* ocena istnienia trendu w oparciu o autokorelację szeregu danych
* ocena zdolności parametru w oparciu o wskaźniki takie jak Cp, czy Cpk

Oprogramowanie musi dać możliwość analizy statystycznej badań klinicznych równoważności dwóch leków lub substancji, lub ogólnie dwóch sposobów leczenia. Analizę można przeprowadzić dla 9 różnych układów badań równoważności (R – lek referencyjny, T – lek testowy):

- R, T

- RT, TR (układ naprzemienny prosty)

- RTR, TRT

- RTTR, TRRT

- RTRT, TRTR

- RRT, RTR, TRR

- RTT, TRR

- RRT, TTR

- RRTT, RTTR, TRRT, TTRR

Oprogramowanie musi dać możliwość wykonania analizy statystycznej od początku do końca, tzn. od fazy planowania aż po analizę otrzymanych danych, w tym:

* wykonanie potrzebnych obliczeń, m.in. oszacowanie wielkości grup oraz analizę mocy statystycznej,
* pozwala utworzyć losowy przydział do grup,
* sprawdza, czy dane spełniają standardowe założenia, które są zazwyczaj wymagane w badaniach,
* wykonuje kluczową analizę z testem statystycznym wskazującym na równoważność lub jej brak.

Program ma zapewniać efektywność analizy statystycznej badań klinicznych oraz zgodność z rekomendacjami organów rejestracyjnych (EMA, FDA).

**Dokument musi być podpisany kwalifikowanym podpisem elektronicznym lub podpisem zaufanym lub podpisem osobistym.**