



PROCEED¹ – MODELOWANIE, OPTIMALIZACJA I SYMULACJA ZŁOŻONYCH PROCESÓW PRODUKCYJNYCH

Michał Iwaniec

StatSoft Polska Sp. z o.o.

Rozwój przedsiębiorstwa jest konieczny dla zachowania konkurencyjnej pozycji na rynku. Już prawie sto lat temu zaczęto stosować narzędzia analityczne do poprawy jakości procesów produkcyjnych. Narzędzia te pozwoliły stosującym je przedsiębiorstwom uniknąć stagnacji i poprawić swoją konkurencyjność. Wkrótce okazało się to niewystarczające, gdyż poprawa jakości odnosiła się tylko do wytwarzanych produktów, a nie do działania całych przedsiębiorstw. Zaproponowano więc podejście typu TQM (*Total Quality Management*), które miało usprawnić pracę całego przedsiębiorstwa, wprowadzić do organizacji pewną kulturę. Wiele przedsiębiorstw po wprowadzeniu tej techniki weszło na wyższy poziom w dziedzinie poprawy i utrzymania jakości, ale mimo wszystko w firmie Motorola wyewoluował nowy nurt nazwany *Six Sigma*, który zapoczątkował pewien przełom.

Six Sigma niewątpliwie zawdzięcza swój sukces nie tylko usystematyzowaniu zarządzania jakością na poziomie przedsiębiorstwa, ale głównie dzięki temu, że zwraca uwagę na potrzebę zbierania danych i wyciągania z nich użytecznej informacji, przy użyciu najróżniejszych metod statystycznych. Nazwa tej metody nie jest przypadkowa i jest odwołaniem do wszechobecnego zastosowania metod analitycznych (statystycznych) stosowanych do obróbki danych, na co *Six Sigma* kładzie główny nacisk i co z pewnością jest źródłem jej sukcesu.

Kiedy strategia *Six Sigma* została rozpowszechniona, okazało się, że w przypadku niektórych firm jest ona ciągle niewystarczająca. Ciągłe istnieją pewne dziedziny, gdzie samo zastosowanie metod statystycznych nie jest wystarczające. Odnosi się to przede wszystkim do zakładów, które wytwarzają zaawansowane technologicznie i drogie produkty. W takim przypadku, nawet jeżeli w firmie wprowadzona została kultura *Six Sigma* i zbierane są odpowiednie dane, to często pojawiają się problemy z ich interpretacją i wyciągnięciem użytecznych informacji. Ponadto pojawiają się problemy organizacyjne, które powodują, że wdrożenie często zaawansowanych metod analitycznych jest trudne. Inżynierowie mają bowiem podejmować kluczowe decyzje na podstawie wyników analiz statystycznych, z czym często są nieobcyi (i trudno takiego obycia od nich wymagać, gdyż statystyka nie jest ich zasadniczą strefą działania). Metoda rozwiązywania tego typu

¹ PROCEED jest znakiem towarowym zastrzeżonym przez firmę Caterpillar Inc, <http://proceed.statsoft.com>.



złożonych zagadnień zawarta jest w nowym produkcie Caterpillar oraz Statsoft – systemie PROCEED. Produkt ten zawiera implementację opatentowanego przez firmę Caterpillar sposobu na tworzenie modeli procesów produkcyjnych oraz wdrażania ich w przedsiębiorstwie. System jest rozwiązaniem wyjątkowym i unikalnym, pozwalającym radzić sobie z problemami przerastającymi strategię *Six Sigma*, głównie dzięki temu, że została zawarta w nim wiedza i wieloletnie doświadczenie firm z sektora produkcyjnego oraz z sektora analizy danych.

PROCEED - przegląd

PROCEED jest gotowym rozwiązaniem dla przedsiębiorstw, które pozwala poradzić sobie z bardzo skomplikowanymi zagadnieniami produkcyjnymi. Został on specjalnie zaprojektowany, aby skutecznie przewidywać zachowania procesów produkcyjnych poprzez odpowiednie ich modelowanie. PROCEED jest w stanie wyodrębnić zależności pomiędzy produktami a procesem ich wytwarzania w oparciu o dane, które zostały już zebrane i są przechowywane w bazach danych.

System pozwala przede wszystkim wdrożyć pewien wypracowany przez firmę Caterpillar sposób postępowania, stosowany podczas modelowania procesów produkcyjnych. Sposób ten został wielokrotnie sprawdzony i wykazał swoją wysoką skuteczność. Główną zaletą PROCEED jest łatwość przeprowadzania całego procesu modelowania, od początku do końca. Program prowadzi użytkownika przez kolejne kroki i nie ma możliwości przejścia do kolejnego etapu, jeżeli poprzednie nie zakończyły się powodzeniem. Tak jak wspomniano we wstępie, system jest przeznaczony dla przedsiębiorstw „dojrzałych” organizacyjnie i informatycznie. Bierze się to stąd, że warunkiem koniecznym do stworzenia dobrych modeli procesów produkcyjnych jest posiadanie odpowiednich zbiorów danych. Między innymi dlatego system jest przeznaczony dla przedsiębiorstw *Six Sigma*, gdyż z założenia powinny one zbierać odpowiednie dane.

Program szczególnie dobrze sprawdza się w przypadku, gdy:

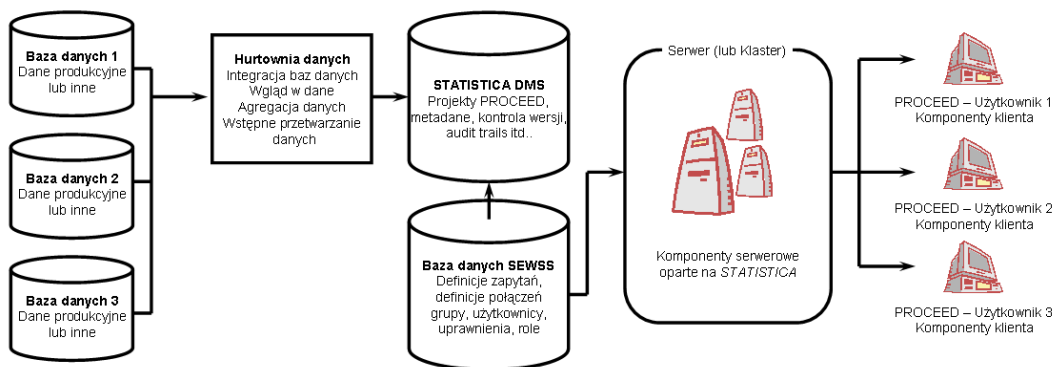
- ◆ procesy produkcyjne są skomplikowane i składają się z wielu etapów,
- ◆ pojawiają się problemy z modelowaniem procesu produkcyjnego,
- ◆ produkt musi spełnić jednocześnie kilka wykluczających się wzajemnie żądań, np. dotyczących mocy silnika i zużycia paliwa,
- ◆ tradycyjne metody analityczne, takie jak planowanie doświadczeń, nie dają się zastosować lub zastosowanie ich jest niemożliwe ze względów praktycznych.

Główną zaletą wdrożenia systemu PROCEED, patrząc pod kątem całego przedsiębiorstwa, jest znaczny zwrot kosztów z inwestycji (*ROI – Return on Investment*), zwłaszcza w dłuższej perspektywie. Niemniej należy zwrócić uwagę na fakt, że rozwiązanie to pozwala osiągnąć dobre wskaźniki ekonomiczne w stosunkowo krótkim czasie, dzięki np. zredukowaniu liczby testów danego produktu, redukcji czasu dostosowywania produktu po zakończeniu produkcji, rozluźnienie wymagań odnoszących się do surowców, jak i granic



specyfikacji parametrów procesu produkcyjnego, redukcji napraw w trakcie trwania gwarancji, czy też poprzez redukcję strat materiału i poprawę jakości produktów. Można wręcz powiedzieć, że oszczędności to główny powód, dla którego system PROCEED został stworzony.

Program oprócz tego, że pozwala na skuteczne modelowanie procesów, kontroluje także przeprowadzane projekty i zapewnia sprawną dystrybucję wiedzy w przedsiębiorstwie. Jest to drugi aspekt związany z kulturą *Six Sigma*. Zapewnione jest to dzięki integracji kilku produktów firmy StatSoft: *STATISTICA Document Management System*, *WebSEWSS* oraz stworzonego od podstaw programu PROCEED. Poniżej przedstawiono architekturę systemu:



Rys. 1. Architektura systemu PROCEED

Dzięki takiej architekturze system zapewnia szerokie możliwości współpracy wszystkich elementów, pracy grupowej oraz zapewnienia bezpieczeństwa danych, jak również stworzonych modeli.

Jednym ze składników systemu jest *WebSEWSS (WebSTATISTICA Enterprise-Wide SPC/Statistical System)*, którego pierwotnym zadaniem było statystyczne sterowanie procesami na poziomie całego przedsiębiorstwa, ale szybko okazało się, że jest to elastyczna i wydajna platforma analityczna i może być wykorzystywana także w dziedzinach niezwiązanych z produkcją. Pozwala na definiowanie połączeń do różnych baz danych i wykonywania automatycznych analiz (monitorów) w oparciu o te dane. Dzięki temu, że jego motorem analitycznym jest program *STATISTICA*, nie musi się on ograniczać tylko do roli systemu SPC, ale może być po prostu centralną platformą analityczną dla całego przedsiębiorstwa. *WebSEWSS* zapewnia dostęp poprzez przeglądarkę internetową do zdefiniowanych wcześniej monitorów, dzięki czemu wiedza w łatwy sposób rozprzestrzenia się w przedsiębiorstwie.

Zarówno pliki, które zostały utworzone za pomocą programu PROCEED oraz *SEWSS*, są przechowywane w systemie *STATISTICA Document Management System (S-DMS)*. Zapewnia on wydajne i bezpieczne zarządzanie różnego rodzaju dokumentami w postaci elektronicznej na poziomie całego przedsiębiorstwa. Wszystkie dokumenty, które



generowane są przez system (np. pliki danych, definicje modeli) znajdują się w systemie *S-DMS*. System daje możliwość zatwierdzania dokumentów, co daje gwarancję, że dokumenty znajdujące się w systemie są poprawne. W przypadku plików, które zawierają definicję modeli, zapewnia to ich poprawność, gdyż do fazy wdrożenia może przejść tylko zwalidowany i zatwierdzony model.

PROCEED - metodyka

Metodyka zawarta w PROCEED stanowi dobrze zdefiniowaną oraz sprawdzoną strategię dla:

- ◆ budowy modeli analitycznych reprezentujących realne procesy i pozwalających na ich symulowanie,
- ◆ przekazania uzyskanych z modelu informacji do odpowiednich osób,
- ◆ weryfikacji i monitorowania wiarygodności stworzonych wcześniej modeli (wygasanie modeli),
- ◆ zarządzania wiedzą (modelami) i wynikami poprzez bezpieczne rozwiązanie zarządzania dokumentami.

W pewnym sensie metodyka PROCEED kieruje myśl zapożyczona z medycyny, a pochodząca od Hipokratesa, „Po pierwsze nie szkodzić”. Możliwość wykrycia warunków, w których na podstawie danych nie da się uzyskać pewnych i dokładnych modeli predykcyjnych, jest tak samo ważna jak sama możliwość budowy modeli. Innymi słowy, czasami dane zebrane w celu opisanie procesu nie mają wystarczającej jakości lub po prostu nie zawierają pomiarów parametrów, które mają najistotniejszy wpływ na interesujące producenta wyjścia procesu. W szczególności może zaistnieć niebezpieczna sytuacja, gdy istnieje model, który jest poprawny, ale daje niepoprawne wyniki dla zakresów wartości, które nie zostały uwzględnione podczas tworzenia modelu. Biorąc ten fakt pod uwagę, PROCEED nie dopuszcza do zaistnienia takiej sytuacji, powiadamiając o tym fakcie użytkownika. W takim przypadku model powinien zostać uzupełniony o nowe zakresy danych wejściowych.

Sercem PROCEED jest przepis na tworzenie modelu, który jest poniekąd podobny do tworzenia modelu w zagadnieniach *data mining*. Istnieje jednak pewna istotna różnica pomiędzy podejściem *data mining* a podejściem wypracowanym przez Catrpillar. Otóż, tak jak wspomniano wcześniej, modelowanie procesów jest zagadnieniem stosunkowo trudnym i wymaga zarówno wiedzy, jak i doświadczenia. Podczas tworzenia projektu *data mining* analityk sam musi zadbać o to, aby wynik modelowania był poprawny, a następnie ktoś musi zadbać o jego wdrożenie i odpowiednie wykorzystanie. Gotowe modele są wykorzystywane głównie przez inżynierów zajmujących się procesem produkcyjnym i ich znajomość narzędzi analitycznych zwykle jest mała. W związku z tym przepis zaproponowany w PROCEED kładzie duży nacisk na wdrożenie modelu, gdyż tak naprawdę to ostatecznie decyduje o sukcesie – możliwość wyciągnięcia z danych użytecznej wiedzy poprzez model. PROCEED proponuje następujące etapy procesu modelowania:

- ◆ Definiowanie danych



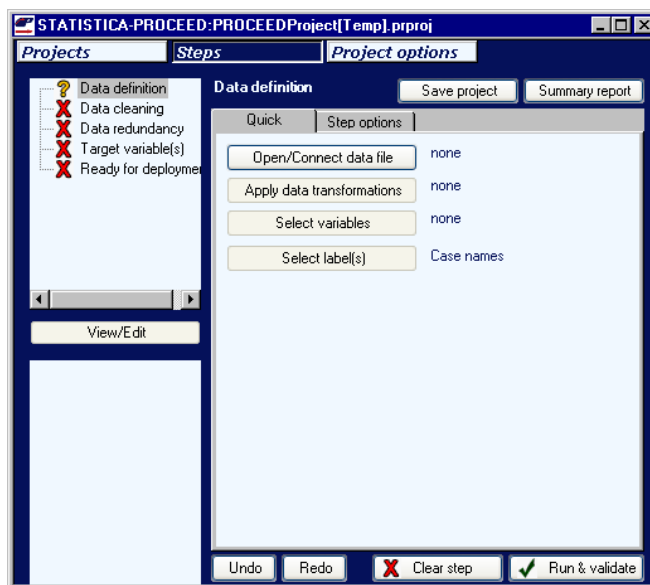
- ◆ Czyszczenie danych
- ◆ Usuwanie redundancji zmiennych
- ◆ Redukcja wymiarów
- ◆ Budowa modelu
- ◆ Symulacja
- ◆ Optymalizacja

Kroki te zostaną dokładniej opisane poniżej.

Definiowanie danych

W pierwszym etapie procesu tworzenia modelu należy oczywiście zdefiniować odpowiedni plik z danymi lub połączenie do bazy danych. W oparciu o te dane, w kolejnych krokach budowany będzie model danego procesu.

Tworzone są zatem zapytania, które wydobędą z baz danych wszystkie istotne dane. Następnie wykonywane są niezbędne przekształcenia oraz wstępne czyszczenie danych. Może ono polegać na usunięciu rekordów zawierających błędne wpisy, czy też przeprowadzeniu wstępnej transformacji zmiennych. W tej fazie nieodzowny wydaje się kontakt analityka (osoby, która będzie budować model) z administratorem bazy danych, w celu uzyskania informacji na temat ich struktury i uzyskania dostępu do odpowiednich danych. Kolejną osobą, która może pomóc analitykowi na tym etapie, jest inżynier, który zna proces i może, opierając się na własnym doświadczeniu, wskazać zmienne procesowe, które w sposób istotny wpływają na jego wyjścia. Informacja ta będzie szczególnie ważna w kolejnych etapach, ze względu na fakt, że takie zmienne nie powinny zostać wyeliminowane.



Rys. 2. Pierwszy krok budowy modelu w PROCEED



Na rysunku 2 pokazane jest standardowe okno programu PROCEED. Z lewej strony widać definicję poszczególnych etapów. Pierwszy etap oznaczony jest znakiem zapytania, który oznacza, że nie został on jeszcze zakończony. Kolejne etapy oznaczone są czerwonymi krzyżykami, co informuje użytkownika o tym, że nie ma możliwości przejścia do tych etapów. Po poprawnym zakończeniu każdego z etapów użytkownik musi zatwierdzić dany etap i wtedy pojawi się możliwość przejścia do kolejnych etapów procesu modelowania. Dzięki takiemu mechanizmowi analityk podąża ściśle określoną ścieżką i prawdopodobieństwo przeoczenia któregoś z wymaganych działań jest minimalne.

Czyszczenie danych

W gruncie rzeczy ten etap jest rozszerzeniem poprzedniego. Dane zazwyczaj będą wymagać dodatkowych przekształceń, takich jak transformacja zmiennych czy też czyszczenie danych w poszczególnych zmiennych. W szczególności zmienne mogą zawierać braki danych, co może niekorzystnie wpłynąć na model. W takim przypadku można np. wypełnić braki danych średnimi dla każdej zmiennej lub usunąć przypadki, w których występują braki danych. Podczas transformacji danych można również przy użyciu kodu *SVB* (*STATISTICA Visual Basic*) stworzyć dowolną transformację i zastosować ją do pól zawierających braki danych. Na tym etapie analityk musi zdecydować, które zmienne z pliku danych będą reprezentować wyjścia procesu, a które wejścia (zmienne te mogą być zarówno ilościowe, jak i jakościowe), oraz musi zdefiniować podzbiór danych, który będzie używany do weryfikacji poprawności stworzonych modeli. Przykładowo jeżeli naszym zadaniem jest znalezienie odpowiedniego modelu do przewidywania dwóch wartości reprezentujących osiągi silnika lotniczego (np. w procesie projektowania), takich jak osiągi podczas wznoszenia oraz podczas normalnego lotu, to te dwa parametry będą zmiennymi wyjściowymi. Natomiast wszystkie mierzone właściwości, które mogą wpłynąć na zachowanie wyjść – typ śmigła, moc silnika itp. – będą zmiennymi wejściowymi.

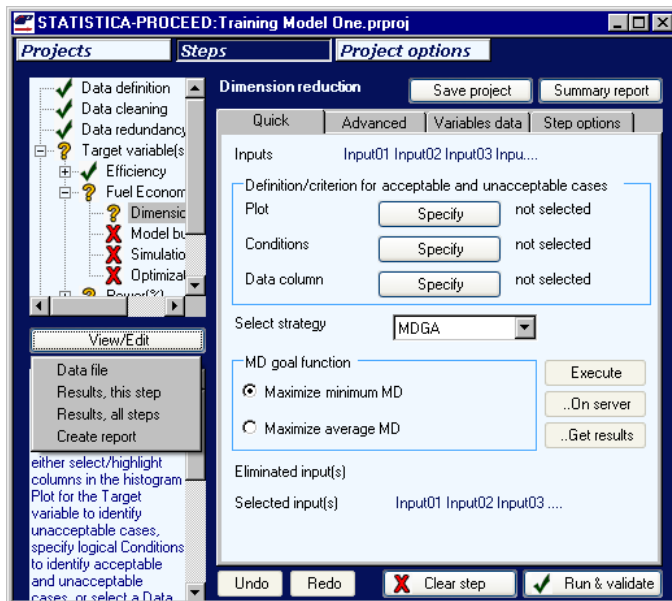
Usuwanie redundancji zmiennych

Ten krok pozwoli na zweryfikowanie, które zmienne w analizowanym pliku danych powtarzają się lub są od siebie mocno zależne. Program automatycznie wyszczególni zmienne, których korelacja jest wyższa od zadanego progu, i zaproponuje listę zmiennych do usunięcia, przy czym użytkownik może zdecydować, czy rzeczywiście chce je usunąć. Przykładem zmiennych, które w całości (choć nie dosłownie) się pokrywają, może być waga mierzona w dwóch jednostkach miary: w kilogramach i w funtach. Do wykrycia tego typu zmiennych zaproponowany został między innymi współczynnik korelacji Pearsona. Można powiedzieć, że wyznacza ona poziom, przy którym możemy stwierdzić, że dwie zmienne są do siebie „proporcjonalne”. Wartość współczynnika korelacji nie zależy od jednostek, które zostały użyte. Dzięki temu wartość korelacji pomiędzy zmiennymi opisującymi przebieg ciśnienia i temperatury w zamkniętym zbiorniku będzie taka sama niezależnie od tego, czy te dwie wielkości wyrażone są w milimetrach słupa rtęci i stopniach Fahrenheita, czy też w PSI i stopniach Celsjusza. Proporcjonalność oznacza w tym przypadku zależność liniową, czyli korelacja jest wysoka, jeżeli zależność może zostać przybliżona linią prostą.



Redukcja wymiarów

Celem tego etapu jest zredukowanie wymiarów zagadnienia predykcyjnego, to znaczy wybranie takiego podzbioru zmiennych wejściowych, dla których zachodzi największe prawdopodobieństwo, że mają wpływ na zmienne wyjściowe, i co za tym idzie, dają możliwość zbudowania dobrego i użytecznego modelu predykcyjnego. Podejście zastosowane na tym etapie, w technice *data mining* określane jest mianem *Doboru i Eliminacji Zmiennych*.



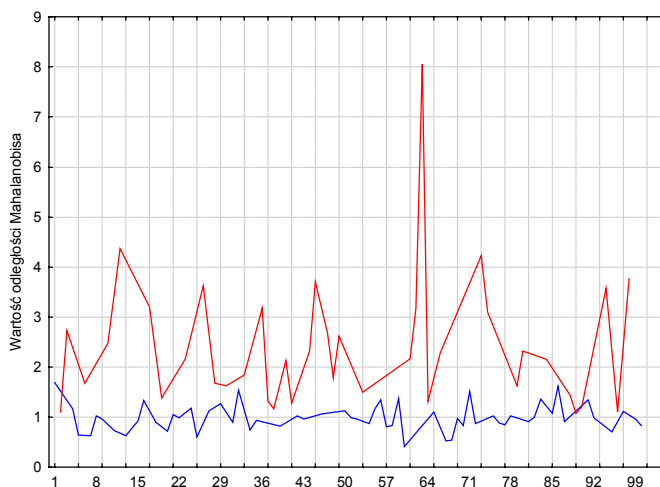
Rys. 3. Okno programu PROCEED – redukcja wymiarów

PROCEED stosuje innowacyjną technikę w celu identyfikacji tychże istotnych zmiennych. Zastosowana tu procedura jest szczególnie efektywna w przypadku, kiedy dane zawierają więcej zmiennych wejściowych niż obserwacji. Nawet jeżeli pewna liczba zmiennych została usunięta w poprzednim kroku, zazwyczaj pozostaje jeszcze duża liczba zmiennych wejściowych, na podstawie których można zbudować model, przy czym część z nich może okazać się nieistotna. Istnieje wiele algorytmów, które pozwalają na rozpoznanie wpływu zmiennych wejściowych na grupę zmiennych wyjściowych. Niemniej dużym wyzwaniem z analitycznego punktu widzenia jest odnalezienie interakcji pomiędzy zmiennymi wejściowymi, które wpływają na wyjścia procesu. Ponadto, proste monotoniczne zależności są zazwyczaj znane wcześniej lub też osoby, które znają proces, takich zależności się spodziewają. Interakcje pomiędzy zmiennymi nie mogą zostać wyrażone w prosty sposób ze względu na dużą złożoność obliczeniową tego zagadnienia. Jeżeli rozważamy około 80 zmiennych wejściowych, to liczba interakcji drugiego rzędu sięga już 3000, a może okazać się, że istotne będą interakcje nawet czwartego rzędu. Zaimplementowana w PROCEED technika łączy wyznaczenie odległości w przestrzeni Mahalanobisa z algorytmem genetycznym (GA – *Genetic-Algorithm*), będącym procedurą optymalizacyjną. Algorytm genetyczny używany jest do maksymalizacji odległości Mahalanobisa pomiędzy



obserwacjami „złymi” a obserwacjami „dobrymi”. Innymi słowy, celem jest znalezienie takiego zestawu wejść, dla których odległość (w sensie Mahalanobisa) pomiędzy przypadkami złymi i dobrymi będzie jak największa. Jasne jest, że dobrane w ten sposób zmienne będą w stanie wskazać taki zestaw danych wejściowych, dla którego otrzymamy (z odpowiednio wysokim prawdopodobieństwem) zadowalające nas wartości wyjściowe. Podejście to wymaga dużo mniejszego nakładu pracy analityka niż w przypadku innych metod optymalizacji i jest w stanie uwzględnić nieliniowe związki pomiędzy zmiennymi, jak również interakcje wysokiego rzędu.

PROCEED daje możliwość przeniesienia tych skomplikowanych i zazwyczaj czasochłonnych obliczeń na dedykowany serwer. Postęp obliczeń można monitorować zdalnie, a po ich zakończeniu wynik zostaje pobrany na stację roboczą. Algorytmy genetyczne są czasochłonne, gdyż z technicznego punktu widzenia nie są procedurami analitycznymi, gdzie optymalne rozwiązanie obliczane jest w oparciu o pewien ustalony wzór. Stosuje się je do zagadnień, gdzie nie istnieją algorytmy o wielomianowej złożoności obliczeniowej, lub mówiąc inaczej, doskonale nadają się do zagadnień, gdzie mamy do czynienia z eksplozją możliwych rozwiązań wraz ze wzrostem liczby zmiennych wejściowych. Należy być świadomym tego, że algorytm genetyczny nie daje pewności odnalezienia optymalnego rozwiązania. Niemniej jednak biorąc pod uwagę fakt, że przeglądnięcie wszystkich możliwych rozwiązań zajęłoby np. tydzień obliczeń, a znalezienie przybliżonego rozwiązania przez algorytm genetyczny zajmuje np. 2 godziny, to zysk jest oczywisty. Poniżej znajduje się wykres, na którym przedstawiona jest odległość pomiędzy przypadkami złymi a dobrymi. Linia, która znajduje się wyżej, oznacza przypadki, których nie można zaakceptować, natomiast linia znajdująca się poniżej przypadki dobre. Widać, że dobrany tutaj zestaw zmiennych wejściowych jest w stanie „odróżnić”, kiedy dany przypadek będzie dobry, a kiedy zły, gdyż pomiędzy liniami utworzony został obszar.



Rys. 4. Wykres odległości Mahalanobisa



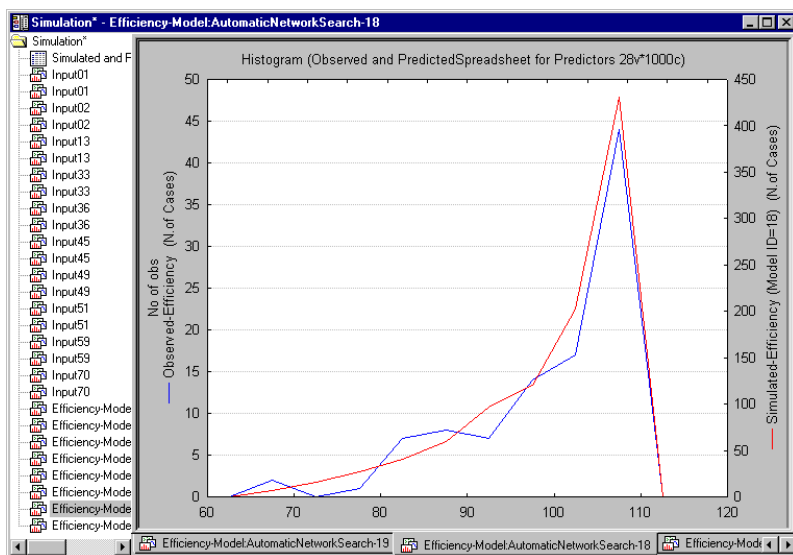
Budowa modelu

Teraz analityk przechodzi do właściwej fazy budowania modelu. Tradycyjnie budowa modeli predykcyjnych często leży w zakresie technik *data mining* lub tradycyjnej statystyki. W przypadku opisywanego tutaj podejścia celem projektantów systemu było maksymalne uproszczenie i automatyzacja budowy modeli, które z odpowiednio dużym prawdopodobieństwem mogą odzwierciedlić rzeczywisty proces. Program automatycznie sprawdzi dużą liczbę modeli opartych na sieciach neuronowych i również automatycznie wybierze te, które są odpowiednio dobrej jakości, aby użyć ich w kolejnych krokach. Używając sieci neuronowych, należy pamiętać o tym, że nie jest to technika oparta na estymacji statystycznej i nie daje najlepszego modelu, jaki można znaleźć, podobnie jak algorytmy genetyczne. W praktyce zostanie znalezionych zazwyczaj kilka podobnych do siebie, odpowiednio dobrych modeli. W wielu przypadkach program jest w stanie automatycznie dobrać odpowiedni model przy standardowych ustawieniach. Jeżeli problem okaże się trudny, użytkownik ma możliwość zmiany ustawień, np. poprzez zwiększenie stopnia komplikacji sieci czy też poprzez zwiększenie liczby sieci, które będą przeglądane. Podobnie jak poprzednio, przeprowadzane w tym kroku obliczenia bywają czasochłonne, w związku z czym program daje możliwość przeniesienia ich na dedykowany serwer, skąd wyniki obliczeń mogą być pobrane po zakończeniu obliczeń. Ponadto analityk ma dostęp do dużej liczby narzędzi graficznych, które pozwalają analizować jakość otrzymanych modeli. Warto zwrócić uwagę na fakt, że interfejs użytkownika został tak zaprojektowany, aby inżynierowie, którzy zajmują się procesami, a niekoniecznie są statystykami lub specjalistami z dziedziny *data mining*, mogli szybko zbudować dobrej jakości modele.

Symulacja

Tradycyjne metody budowy modeli zazwyczaj kończą się na opisanym w poprzednim punkcie etapie. Po dobraniu odpowiedniego modelu, co zostało potwierdzone odpowiednimi statystycznymi kryteriami do oceny jego dobroci, model jest akceptowany i wdrażany. Podczas wielu lat tworzenia i wdrażania modeli przy użyciu technik *data mining*, w celu późniejszego uzyskania z nich informacji pozwalających na zastosowanie ich w praktyce stało się jasne, że tradycyjne kryteria często są niewystarczające. Przypomnieć trzeba, że w tym przypadku celem projektowania jest ostateczne wdrożenie modelu, który ma być pomocny między innymi w przeprowadzaniu scenariuszy „what-if?” („co jeśli”) i w projektowaniu zaawansowanych technologicznie, a co za tym idzie drogich urządzeń. Dlatego nie jest wystarczające, aby oceniać dokładność modelu tylko w sensie statystycznym (np. stopień korelacji pomiędzy wartościami przewidywanymi i obserwowanymi). Istnieje potrzeba, aby model dokładnie odzwierciedlał rozkład właściwości wyjściowych procesu. Rozważmy sytuację, gdzie rozkład wartości wyjściowych jest silnie skośny. Co będzie, jeżeli z modelu sieci neuronowych uzyskamy predykcje, które będą mocno skorelowane z obserwowanymi wartościami, ale będą podlegać rozkładowi normalnemu? W takim przypadku pomimo tego, że model będzie poprawny ze statystycznego punktu widzenia, jest naturalne, że w pewien sposób będzie on tracił część informacji o rzeczywistym działaniu procesu. Z tego właśnie powodu PROCEED wymaga od użytkowników jeszcze jednego kroku, który pozwoli ocenić dobroć dopasowania modelu właśnie pod tym kątem.

Do zmiennych wejściowych dopasowywane są teoretyczne rozkłady i stosowany jest algorytm, który pozwoli dokładnie odtworzyć korelacje pomiędzy zmiennymi. Na tej podstawie generowane są dane, które wchodzi do modelu, w wyniku czego na jego wyjściach otrzymywane są modelowane rozkłady. Jeżeli rozkłady te są takie same jak te otrzymane w oparciu o dane rzeczywiste, wtedy możemy mieć większą pewność, że dany model uchwycił leżącą głębiej dynamikę procesu. Takie podejście, oparte na symulowaniu wejść jest unikalne i jednocześnie wiele razy udowodniło już swoją siłę. Metoda ta jest również bezcenna w przypadku, kiedy mamy do czynienia z bardzo małymi zbiorami danych (kilka przypadków i brak zbioru testowego). W takim przypadku wybrany zostaje model, który najlepiej zachowuje się podczas symulacji.



Rys. 5. Przykładowe porównanie rozkładu otrzymanego na podstawie danych i modelu

W tym kroku otrzymujemy już gotowy do wdrożenia model, który będzie reprezentował badany proces. Stworzone tu modele mogą zostać użyte do predykcji w sterowaniu jakością lub do podjęcia decyzji o poddaniu wyprodukowanego elementu fizycznemu testom.

Optymalizacja

Wykonywanie tej części nie jest konieczne, ale zazwyczaj jest użyteczne. W poprzednim kroku zostały zdefiniowane modele, które opisują, w jaki sposób wejścia procesu wpływają na jego wyjścia, w związku z czym mogą one teraz zostać użyte do takiego doboru wartości wejściowych, aby wartości wyjść procesu były jak najbardziej zbliżone do optymalnych. PROCEED daje możliwość minimalizacji lub maksymalizacji wartości wyjściowych, a także optymalizacji wskaźników zdolności (C_{pk}) dla interesujących nas wyjść procesu. Analityk ma możliwość określenia zakresu zmiennych wejściowych, poza które procedura optymalizacji nie może wychodzić. Analityk może narzucić również inne ograniczenia, co spowoduje, że możliwość zmiany niektórych wejść może zostać



zablokowana, np. „z 20 dostępnych wejść modelu, które wpływają na wyjścia, znajdź 4, które mają największy wpływ i przy ich użyciu znajdź optymalną wartość wyjścia”. Zauważmy, że nie zostaną znalezione tylko najlepsze wartości zmiennych wejściowych, które generują optymalne wartości wyjść, ale także podane zostaną parametry funkcji rozkładów, które będą maksymalizować, np. średnią wartość zmiennej wyjściowej. W związku z tym przeprowadzana tutaj optymalizacja nie jest zwykłą optymalizacją punktową, ale optymalizacją parametrów rozkładów zmiennych wejściowych. Jest to sprawą kluczową w zagadnieniach produkcyjnych, gdyż odzwierciedla rzeczywistą naturę fizycznych procesów, których naturalną własnością jest zmienność. Okazuje się, że w zagadnieniach produkcyjnych nierealne jest, aby ustawić wartości wejściowe fizycznego procesu na konkretnym poziomie, a następnie wartości te utrzymać. W większości przypadków rozkłady zmiennych wejściowych mogą zostać w pewnym stopniu zmodyfikowane, np. poprzez odrzucenie zakupionych elementów, których parametry znajdują się poniżej pewnego krytycznego progu, nawet jeżeli spełniają one narzucone wcześniej specyfikacje.

Ponadto dzięki temu, że po zoptymalizowaniu otrzymaliśmy rozkłady zmiennych wejściowych zamiast pojedynczych (optymalnych) wartości (punktów), nierzadko okazuje się, że wartości zmiennych wejściowych mogą zmieniać się w szerokim zakresie, nie powodując przy tym istotnych zmian rozkładów krytycznych wartości wyjściowych procesu. Taka wiedza okazuje się zazwyczaj wartościowa ze względu na fakt, że szersze tolerancje dla wejść przekładają się zazwyczaj na niższe koszty (np. można wybrać tańszego dostawcę).

Zmienne wyjściowe mogą być optymalizowane pojedynczo, ale zazwyczaj istnieje potrzeba, aby zoptymalizować kilka wielkości wyjściowych, których optymalne wartości są sobie przeciwstawne, np. zużycie paliwa i moc silnika – wiadomo, że wraz ze wzrostem mocy silnika będzie rosło zużycie paliwa. Celem będzie zatem otrzymanie takich wartości, które pozwolą uzyskać jak największą moc przy jak najmniejszym zużyciu paliwa. PROCEED przeprowadzi wtedy procedurę, która znajdzie rozwiązanie kompromisowe przy użyciu odpowiedniej funkcji celu.

Wdrożenie modelu – *Actionable Decision Environment* (ADE)

Tak jak wspomniano wcześniej, główną przeszkodą podczas implementacji zaawansowanych metod analitycznych, które mają usprawnić proces, jest kwestia rozpropagowania modelu w przedsiębiorstwie i jego skutecznego stosowania. Aby osiągnąć sukces, istotne spostrzeżenia, które mogą poprawić jakość i wydajność procesu, muszą zostać sprawnie rozpropagowane w przedsiębiorstwie (komunikacja) oraz przez wszystkich odpowiednio zrozumiane. Niezbędne jest wypracowanie metod współpracy pomiędzy analitykami (statystykami) a inżynierami, którzy mają wiedzę na temat funkcjonowania procesu. PROCEED jest rozwiązaniem przeznaczonym dla przedsiębiorstw i pozwala na wdrażanie zaawansowanych metod analitycznych do rutynowych zadań na wszystkich szczeblach przedsiębiorstwa, jak również pozwala zbierać (akumulować) wiedzę o procesach w sposób ciągły, systematyczny i przemyślany. Ponadto system ten nie został stworzony tylko po



to, aby wykonać jeden ważny projekt analityczny. PROCEED należy rozumieć jako proces ciągły, metodę postępowania podobną do szeroko rozpowszechnionej strategii *Six Sigma*. Zawiera on wszystkie narzędzia, które wspomagają wdrożenie metodyki PROCEED w przedsiębiorstwie.

W poprzednim rozdziale opisana została procedura tworzenia samego modelu oraz optymalizacji wartości wyjściowych procesu. PROCEED posiada także elastyczne narzędzia do rozpowszechniania stworzonych w ten sposób modeli w przedsiębiorstwie, dzięki czemu osoby, które będą podejmować ostateczne decyzje (w oparciu o stworzone modele), będą mieć do nich łatwy dostęp. Osoby te mogą przeglądać wyniki generowane przez modele, badać je poprzez ręczne zmiany parametrów wejściowych, modyfikować czy też odrzucać wyniki pochodzące z modelu. ADE jest interfejsem użytkownika opartym na technologii internetowej, dzięki czemu może w łatwy sposób objąć swym zasięgiem całe przedsiębiorstwo. Pozwala on między innymi na przeglądanie wyników, przeprowadzanie scenariuszy „what-if” („co jeśli”) czy też monitorowanie dobroci dopasowania istniejącego modelu do aktualnych danych.

Mówiąc inaczej ADE jest narzędziem, za pomocą którego informacja uzyskana podczas budowy modelu jest przetwarzana na wiedzę, którą można wykorzystać do poprawy procesu. Ponadto wiedza ta jest przedstawiana w sposób najbardziej przystępny dla ostatecznego użytkownika modelu. Biorąc pod uwagę fakt, że PROCEED jest pewną metodyką postępowania, która obejmuje swym zasięgiem całe przedsiębiorstwo, ADC jest motorem, który przenosi informację uzyskaną podczas modelowania do osób, które podejmują decyzje związane z procesami. Dzięki ADE najbardziej skomplikowane procedury analityczne (np. sieci neuronowe) są widziane przez użytkownika jako proste w interakcji „pudełka”. Dzięki temu użyteczność modelu jest większa, bo jest on eksplorowany przez inżynierów, którzy doskonale znają swój proces i mogą teraz łatwo połączyć wiedzę praktyczną z sugestiami pochodzącymi od modelu. ADE składa się z kilku elementów:

ADE Configurations

Są to szablony, w których znajdują się odwołania do danych (lokalnie lub poprzez profile systemu SEWSS) i do stworzonego modelu, mające na celu ułatwienie jego badania. Dostęp do nich jest realizowany za pomocą przeglądarki internetowej i pozwala na natychmiastowy dostęp do informacji uzyskanych z modelu, po nadejściu nowych danych z procesu produkcyjnego. ADE Configurations są tworzone w prosty sposób przy użyciu interfejsu ADE Builder. Odpowiednie elementy, które mają znaleźć się w szablonie wstawiane są po prostu z biblioteki standartowych elementów. Ponadto stworzone szablony mogą następnie być załadowane do systemu SEWSS i działać jako standardowe monitory SEWSS.

ADE Objects

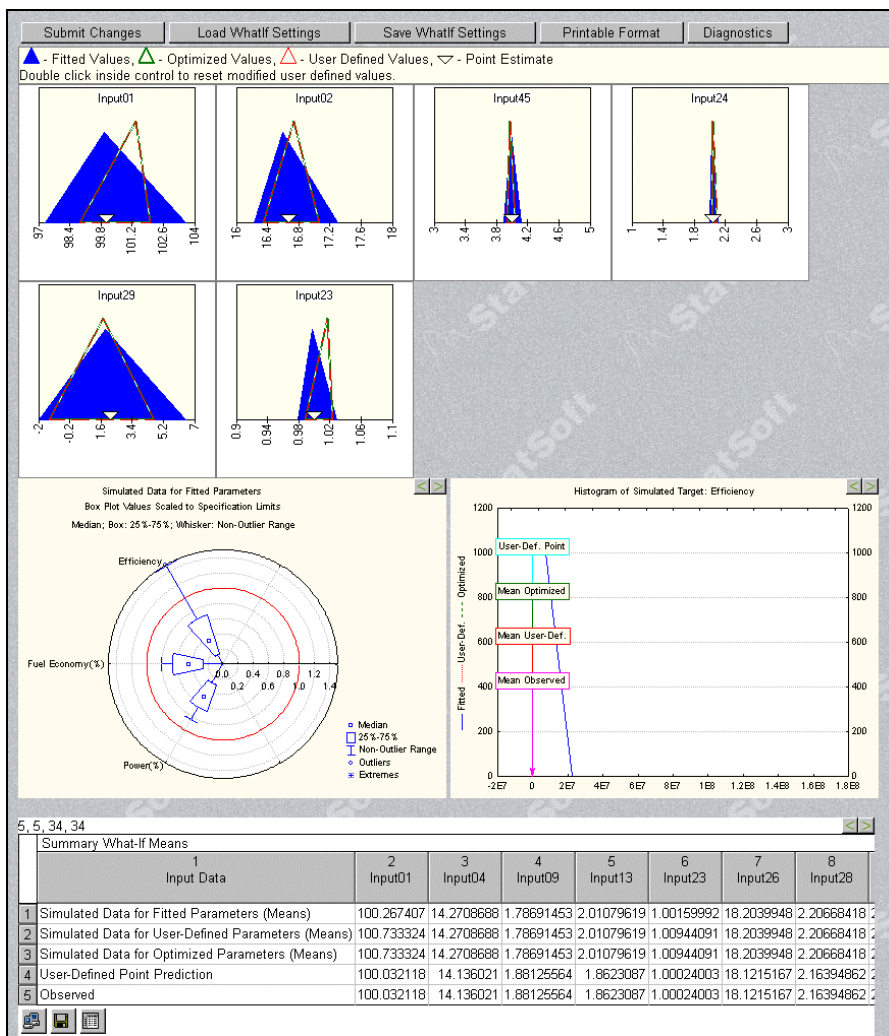
ADE Configurations zawierają w sobie ADE Objects, które następnie mogą być używane w sposób interaktywny. Stanowią one narzędzia, które pozwalają na podłączenie modeli predykcyjnych stworzonych w PROCEED do konkretnych narzędzi graficznych, co zachęca użytkownika do eksploracji modeli i analizowania pochodzących od nich sugestii.



Bardzo dużo elementów ADE Objects odnosi się do narzędzi graficznych oraz prezentacji wyników. Inne natomiast zostały zaprojektowane specjalnie po to, aby monitorować dobroć dopasowania aktualnie używanego modelu, dzięki czemu zmieniające się warunki (np. zakresy wartości zmiennych wejściowych), które mogą spowodować błędne działanie modelu są wykrywane tak szybko, jak to tylko jest możliwe. Obiekty tego typu mogą wręcz nadać modelowi status „niezwalidowanego” i model taki powinien zostać ponownie zweryfikowany, tak aby uwzględnić zmieniające się warunki.

Przykład wdrożenia

Na poniższym rysunku znajduje się przykładowe wdrożenie stworzonego modelu, w środowisku ADE.



Rys. 6. Przykładowy interfejs ADE



Wybranych zostało 6 zmiennych wejściowych (najbardziej istotnych), za pomocą których inżynier może przeprowadzać scenariusze „what-if”. Każda ze zmiennych przedstawiona jest za pomocą rozkładu trójkątnego, który inżynier może zmieniać za pomocą myszki i patrzeć, jak zmieniły się wartości zmiennych wyjściowych. Mamy tu do czynienia z trzema zmiennymi wyjściowymi, których jakość została przedstawiona na wykresie radarowym. Należy pamiętać o tym, że zmienne wyjściowe są sobie przeciwstawne i próba poprawy jednej z nich prawdopodobnie pogorszy dwie pozostałe. Dlatego taki sposób wdrożenia modelu jest szczególnie cenny, gdyż pozwala inżynierowi szybko zorientować się we właściwościach procesu. Wykres radarowy jest również łatwy w interpretacji; im bliżej środka znajdują się umieszczone na nim wykresy ramka-wąsy, tym lepiej. Zauważmy, że na rysunkach rozkładów zmiennych wejściowych znajdują się trzy kształty. Pole oznacza rozkład zaobserwowany w trakcie symulacji, jeden z konturów oznacza optymalną wartość wyznaczoną przez algorytm genetyczny, natomiast drugi kontur może być modyfikowany przez inżyniera.

To, co przede wszystkim zwraca uwagę podczas eksploracji przedstawionego powyżej modelu, to fakt, że zoptymalizowane rozkłady dają dużo lepsze parametry zmiennych wyjściowych niż rozkłady pochodzące z symulacji. Ponadto dzięki temu, że użytkownik może ręcznie „poprawić” dany rozkład, ADE staje się narzędziem, dzięki któremu można szybko wydobyć z modelu często bezcenną wiedzę.

PROCEED – role

W przypadku PROCEED bardzo ważne jest właściwe zdefiniowanie odpowiednich ról w systemie dla poszczególnych użytkowników. W przeciwieństwie do tradycyjnych systemów *data mining*, gdzie rozpowszechnienie metod statystycznych w przedsiębiorstwie jest trudne ze względu na potrzebę angażowania wyspecjalizowanych pracowników, PROCEED będzie obejmował dużo większą grupę ludzi ze względu na prostotę jego użycia.

Kluczową rolę w systemie, który służy przede wszystkim do modelowania, odgrywa „*Model Builder*”. Jest to osoba, której rolą jest budowa modeli analitycznych pozwalających na poprawę jakości procesu produkcyjnego. Zazwyczaj osobie takiej nie będzie obca analiza danych, jak również pewna wiedza z zakresu tworzenia modeli predykcyjnych. W organizacjach, które wdrożyły strategię *Six Sigma*, rolę tę zazwyczaj przejmują *Black Belts* czy też *Master Black Belts*, którzy przeszli szkolenie z zakresu użytkowania PROCEED. Zwykle *Model Builder* będzie pracował zespołowo razem z inżynierami, którzy znają dany proces. Dzięki temu osoba tworząca modele może uzyskać bezcenną empiryczną wiedzę, jak również skupić się na tych zagadnieniach, które w danym momencie są najbardziej istotne. W organizacjach *Six Sigma* nie należy do rzadkości fakt, że zespół prowadzony jest przez *Master Black Belta*, który jednocześnie pełni rolę *Model Buildera* w systemie PROCEED, gdyż zazwyczaj osoba taka posiada dużą zdolność do rozwiązywania problemów. *Model Builder* tworzy modele i następnie zapisuje je w środowisku *STATISTICA Document Management System (SDMS)*. Jest on również odpowiedzialny za ponowne



walidowanie modelu, jeżeli straci on swoją ważność, np. ze względu na zmieniające się zakresy danych. Jego zadaniem jest również nadzór nad wdrażaniem modelu.

Oczywisty jest fakt, że *Model Builder* do stworzenia modelu potrzebuje odpowiednich danych. Za zarządzanie i dostarczenie odpowiednich dla danego problemu danych odpowiedzialny jest kolejny użytkownik: *Database Query Administrator (DQA)*. *SEWSS* dostarcza narzędzi, które pozwalają na połączenie z bazami danych i stworzenie odpowiednich zapytań, które wyekstrahują odpowiednie dane potrzebne do budowy modelu. Architektura *SEWSS* pozwala na dostęp użytkowników do zapytań, nawet jeśli oni sami nie mają odpowiednich uprawnień. Jedynie osoby, które tworzą zapytania, muszą mieć odpowiednie uprawnienia. Zazwyczaj *Database Query Administrator* w porozumieniu z inżynierami czy też użytkownikami końcowymi tworzy odpowiednie szablony zapytań do baz danych, które potem mogą zostać wykorzystane np. podczas budowy modelu. Szablony zapytań mogą być stałe, aby np. pobierać dane z pewnego okresu, lub mogą zezwalać użytkownikom na zawężenie zapytania. Po stworzeniu zapytania przenoszone jest ono do systemu zarządzania dokumentami *S-DMS* w celu zachowania integralności w systemie.

Zazwyczaj rolę *DQA* przejmują administratorzy baz danych w przedsiębiorstwie. W trakcie wdrażania systemu, zanim wszystkie zapytania zostaną zdefiniowane, rola ta wymaga większego nakładu pracy. Natomiast gdy system zostanie już wdrożony, niemal wszystkie niezbędne zapytania są już zbudowane i od tego użytkownika nie jest w czasie normalnej pracy systemu wymagany większy nakład pracy.

Kolejną ważną rolę w systemie *PROCEED* jest *ADE Builder*. Osoba ta jest odpowiedzialna za dostarczenie stworzonego modelu użytkownikom końcowym. Zazwyczaj rola ta jest przejmowana przez te same osoby, które zajmują się budową modeli, ze względu na fakt, że efektywne wdrożenie modelu, co było już wcześniej podkreślane, jest tak samo ważne jak jego stworzenie.

PROCEED w działaniu – podsumowanie

Od momentu, kiedy *PROCEED* zostanie w pełni wdrożony w przedsiębiorstwie, będzie wspierał rozwój metody, w oparciu o którą będzie możliwe opanowanie procesów. Dzięki zastosowaniu zaawansowanych, opartych na wykorzystywaniu wiedzy zawartej w danych metodyki otrzymujemy solidne podstawy do tego, aby w przyszłości stworzyć organizację sterowaną przez realną wiedzę zawartą w danych. W takim przedsiębiorstwie przewidywane możliwości i poglądy muszą być poparte dowodami bazującymi na danych. Można powiedzieć, że podejście to jest rozszerzeniem strategii *Six Sigma* i jego pomyślne stosowanie może przenieść przedsiębiorstwo na wyższy poziom wydajnościowy i jakościowy. Każdy z użytkowników, który uczestniczy w procesie *PROCEED*, ma jasno wyznaczoną rolę i jest wspierany odpowiednim zestawem narzędzi, który pozwoli osiągnąć zamierzony cel. Końcowi użytkownicy dzięki temu, że zdobytą podczas modelowania informację mogą wykorzystać „w akcji”, są w stanie łatwo sprawdzić swoje podejrzenia w stosunku do zachowania procesu, sprawdzić, czy decyzje, które chcą podjąć odnośnie zmian, są trafne i znajdują potwierdzenie w modelu, co w rezultacie powoduje, że proces



jest pod większą kontrolą. Ponadto dzięki temu, że wszystkie dokumenty przechowywane są w systemie zarządzania dokumentami, wiedza o procesach jest akumulowana, zachowywane są wszystkie stare wersje dokumentów oraz modeli i historia ich zmian. Daje to pełny wgląd w działanie systemu. PROCEED jako przepis na tworzenie i wdrażanie modeli jest zawarty w pełnym, zawierającym wszystkie potrzebne elementy środowisku, które oparte jest na architekturze klient-serwer. Dzięki temu może w sposób efektywny dostarczać zaawansowanych metod analitycznych przy zastosowaniu metod typu „work-flow”, co powoduje, że projekty są przeprowadzane sprawnie.