

Marian Krupa
Maciej Pękala

Zastosowanie techniki „Check Alert” w zarządzaniu jakością danych oraz analiz biznesowych w firmie

Wstęp

Użyteczność analiz biznesowych jest wprost proporcjonalna do jakości danych użytych do ich przygotowania. Proces zarządzania danymi biznesowymi zarówno w wymiarze wewnętrznym jak i zewnętrznym jest niezwykle trudny i obciążony różnego typu ryzykami. Przebiega on zwykle w sposób wystandaryzowany, w ramach wcześniej zdefiniowanych i zatwierdzonych procedur czy też algorytmów systemowych funkcjonujących w ramach oprogramowania klasy Business Intelligence. Istotną jednak kwestią nie jest sam fakt sprawnego dostarczenia raportu w odpowiedni sposób i w wymaganym czasie ale również rzetelność wyników, które mają niezwykle istotny wpływ na jakość podejmowanych decyzji w firmie. Jak każdy inny proces, również proces zarządzania danymi (informacją biznesową) wymaga zastosowania odpowiednich metod i narzędzi. Skuteczność w zakresie zapewnienie wysokiej jakości danych poprzez zaplanowane i wdrożone rozwiązania systemowe, determinuje ostatecznie jakość całego systemu zarządzania informacją w każdej firmie.

Celem niniejszego opracowania jest ocena użyteczności autorskiej techniki „Check Alert” w zakresie wsparcia użytkownika kluczowego / właściciela danych w procesie weryfikacji rzetelności analizowanych raportów, generowanych w systemach klasy BI. Referat obejmuje próbę odpowiedzi na następujące pytania: 1) W jaki sposób przebiega proces zarządzania jakością danych w firmie zgodnie z zasadą 8R i 5ZRR? 2) Czym są tzw. „defekty danych” informacji biznesowej przechowywanej w bazach danych? 3) Na czym polega modelowanie symulacyjne i jaka jest jego użyteczność w doskonaleniu procesu zarządzania danymi biznesowymi? 4) Czym jest autorska technika diagnostyczna „Check Alert”? oraz 5) Jak wykorzystać oprogramowanie SAP Analytics Cloud w ocenie jakościowej danych biznesowych z perspektywy oceny rzetelności generowanych raportów?

W celu odpowiedzi na postawione pytania zastosowano metodę analityczno-syntetyczną w zakresie literatury naukowej i branżowej oraz modelowanie symulacyjne zrealizowane przy pomocy oprogramowania klasy BI SAP Analytics Cloud oraz techniki „Check Alert”.

Podsumowaniem przeprowadzonej analizy oraz wykonanej symulacji, w oparciu o dane finansowe spółek giełdowych, jest opracowane w formie syntezy zestawienie wniosków i rekomendacji.

1. System zarządzanie informacją biznesową w firmie zgodnie z zasadą 8R oraz 5ZRR – w kierunku integracji przepływów logistycznych i informacyjnych

Problematyka zarządzania jakością danych, w tym analiz biznesowych, jest omawiana w literaturze przedmiotu m.in. w kontekście analizy niesprawności informacyjnej („choroby informacji”). Powyższe zachowania o charakterze patologicznym są przedstawiane z punktu widzenia zarówno: 1) produktu jako efektu realizowanego procesu zarządzania informacją i komunikacją w firmie jak też 2) procesu, który prowadzi do wytworzenia w/w produktów, jakim są raporty i analizy biznesowe¹. Niesprawność danych, informacji jako produktu obejmuje²:

- Przeciążenia informacyjne – dostarczenie nadmiaru informacji, danych biznesowych co przejawia się wydłużeniem czasu analizy, wzrostem kosztów, ryzykiem niespójności oraz obniżeniem motywacji i zaangażowania się decydenta;
- Dwuznaczność informacji – opracowanie raportu, analizy, która prowadzi do sprzecznych, dwuznacznych wniosków;
- Anemia informacyjna lub osłabienie „poła widzenia” – zestaw niepełnych danych, danych nieaktualnych lub nadmiarze danych nieadekwatnych do rozwiązywanego problemu decyzyjnego;
- Przekłamanie informacyjne – błędne powiązanie kategorii danych z ich zawartością lub też dla tej samej kategorii podaje się dwie różne treści, zestawy danych;
- Błędnie ustrukturyzowane formularze jako nośniki nieczytelnej informacji – skomplikowane struktury, brak logicznego powiązania poszczególnych tabel i pól, nadmiar ręcznie wprowadzanego tekstu, danych, chaotyczne opracowanie edytorskie.

¹ Z. Martyniak, *Podstawy diagnozowania informacji jako produktu i procesu* [w:] Z. Martyniak (red.), *Zarządzanie informacją i komunikacją. Zagadnienia wybrane w świetle studiów i badań empirycznych*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2000, s. 11.

² Tamże, s. 12-17.

Z perspektywy oceny jakościowej procesu zarządzania informacją należy szczególnie zwrócić uwagę na³:

- Zaleganie informacji – pojawianie się tzw. „wąskich gardeł” w procesie tworzenia i przekazywania danych, analiz biznesowych;
- „Dystorsję” informacji – dwuznaczność informacji, danych wynikająca z różnych formuł, standardów jej wprowadzania, przetwarzania i analizy;
- Zawał informacyjny – brak danych, informacji w systemie wynikający z braku wymaganych zasobów, sprawnych procedur lub też braku integracji w zakresie przepływów towarowych, finansowych z przepływami informacyjnymi.

Niezwykle istotnym elementem zapobieganiu tzw. „zawałom informacyjnym” w procesie zarządzania informacją jest zapewnienie właściwej integracji. W przepływach zarówno towarowych (perspektywa logistyczna) jak i informacyjnych (perspektywa decyzyjna) podstawowym wyznacznikiem sukcesu jest wdrożenie systemu zintegrowanego, na przykład w oparciu o koncepcję ośmiu warunków, wymagań (8R) oraz 5 zasad racjonalnego zarządzania (5ZRR).

W pierwszym wypadku koncepcja 8R (*8 Right*) wskazuje na konieczność zapewnienia w ramach systemu zarządzania informacją i komunikacją w firmie: 1) właściwej informacji (*right information*); 2) we właściwej ilości (*right quantity*); 3) o odpowiedniej jakości (*right quality*); 4) we właściwe miejsce (*right place*); 5) w odpowiednim czasie (*right time*); 6) do właściwego odbiorcy, decydenta, pracownika (*right customer*); 7) przy odpowiednim poziomie kosztów (*right cost*) oraz 8) w odpowiedzialny sposób z punktu widzenia środowiska społecznego i naturalnego (*right way*)⁴. Koncepcja 8R definiuje konieczność równoczesnego monitorowania i oceny jakości danych i systemu analiz biznesowych z perspektywy powyżej opisanych perspektyw. Jest to zadanie stosunkowo złożone i wymagające zapewnienia odpowiednich zasobów i metod, w tym zaawansowanych technologii informatycznych.

Z kolei, zgodnie 5ZRR zintegrowany proces zarządzania informacjami oparty o najwyższe standardy jakościowe (zasada 8R) powinien prowadzić do podejmowania decyzji w celu: 1) zapewnienia właściwej równowagi pomiędzy celami ekonomicznymi a kosztami społecznymi/ekologicznymi (zasada optymalizacji); 2) wyeliminowania sprzeczności pomiędzy różnymi aspektami decyzyjnymi (zasada współdziałania) 3) minimalizacji sprzeczności, unikanie konfliktu celów (zasada kompatybilności; 4) podejmowanie decyzji w

³ Tamże, s. 18-20.

⁴ Mangan, Lalwani, Butcher, *Global Logistics and Supply Chain Management*, Wiley & Sons, 2008.

oparciu o sprawdzone i wiarygodne dane (zasada wiarygodności); 5) uwzględnienia uwarunkowań branżowych, technologicznych, prawnych itd.(zasada konkurencyjności)⁵.

Zarówno koncepcja 8R oraz model 5ZRR wyznaczają nam fundamentalne przesłanki prowadzące do wdrożenia i utrzymania wysokiej jakości zarządzania systemem informacji. W dalszej części niniejszego opracowania w sposób bardziej szczegółowy zostanie opisany aspekt procesu podejmowania decyzji biznesowych w firmie w perspektywie występowania tzw. „defektów danych” w administrowaniu bazami danych.

2. „Defekty danych” a jakość podejmowanych decyzji biznesowych w firmie

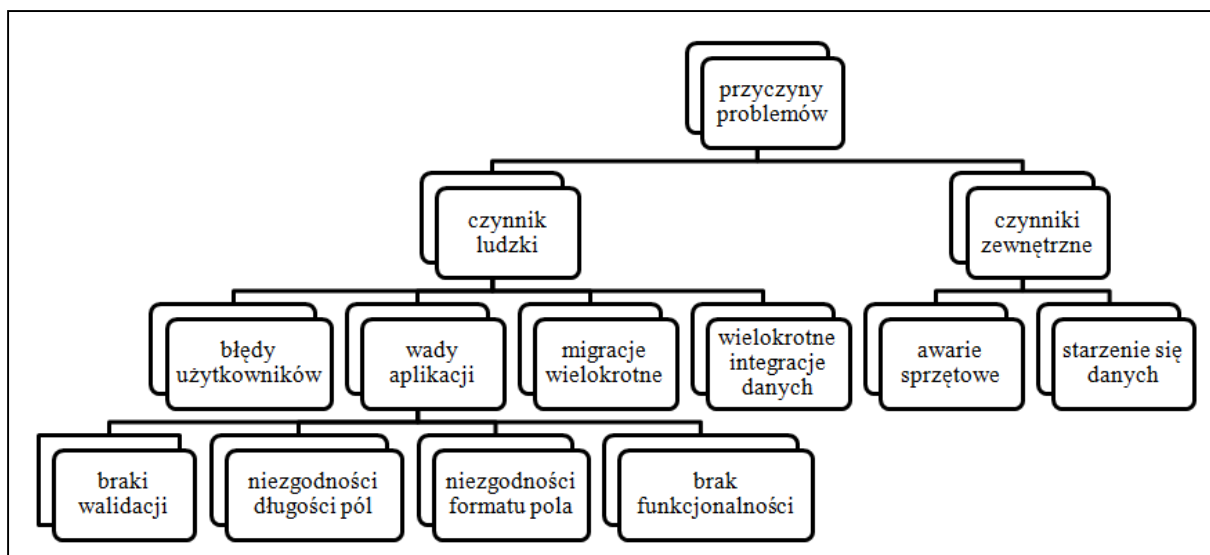
Najogólniej rzecz ujmując, dane tworzące system zarządzania informacjami w firmie, przechowywane są w bazach danych. Z nimi związane są takie właściwości danych jak: współdzielenie, integracja, integralność, bezpieczeństwo, abstrakcja i niezależność⁶. W rozważaniach będzie nas interesować integralność danych, czyli dokładne odzwierciedlenie obszaru analizy, których baza jest modelem. W praktyce oznacza to, że jeżeli istnieją związki w rzeczywistym świecie między obiektami reprezentowanymi przez dane w bazie danych, to zmiany dokonane w rzeczywistości muszą znaleźć odbicie w bazie, a zmiany po jednej stronie takiego związku w bazie, powinny być dokładnie odzwierciedlone po innych stronach związku. Integralność oznacza, że powinniśmy mieć zaufanie do tego co znajduje się w bazie danych w zakresie odpowiedniości między faktami przechowywanymi w bazie danych a rzeczywistością.

To zaufanie może być nadszarpnięte przez tzw. „defekty danych” rozumiane szeroko, a w szczególności jako uszkodzenie, wada lub brak. Ich źródła mogą być rozmaite, a pośrednio jest to klasyfikacja (kategoryzacja) defektów, co uwidoczniło na Rysunku 1⁷.

⁵ A. Wiktorowska-Jasik, *Organizacja procesu transportu ładunków transportem zintegrowanym* [w:] I.N. Semonov (red.), *Zintegrowane łańcuchy transportowe*, Difin, Warszawa 2008, s. 40-41.

⁶ Szczegółowe omówienie tych cech można znaleźć w Paul Beynon-Davies, *Systemy baz danych*

⁷ Dariusz Gałęzowski, Mariusz Sienkiewicz, *Zarządzanie jakością danych*, Warszawa SGH 06.12.2016r, <https://studylibpl.com/doc/684763/jako%C5%9B%C4%87-danych>, dostęp 27.07.2019



Rysunek 1. Źródła powstawania defektów. Opracowano na podstawie prezentacji SGH, Warszawa 06.12.2016 r.⁸.

Janicki⁹ uważa, że cechami decydującymi o jakości bazy danych są: kompletność informacji, poprawność informacji, spójność danych, aktualność oraz dziedzinność. Przybliżmy po krótku te pojęcia:

- kompletność informacji – to stopień wypełnienia rekordów bazy danymi;
- poprawność informacji – oznacza ich zgodność ze stanem faktycznym z rzeczywistością;
- spójność danych – oznacza prawidłowość wzajemnych relacji, prawidłowość powiązań między parametrami, niewykluczanie się danych i odniesienia;
- aktualność – to zgodność ze stanem faktycznym mimo ciągłych zmian rzeczywistości;
- dziedzinność – to obszar informacji do gromadzenia.

Należy się równocześnie zastanowić jakie są przyczyny złej jakości danych. Janicki¹⁰ wymienia:

- brak kompletności na etapie wprowadzania,
- brak poprawności na etapie wprowadzania,
- brak uaktualnień,

⁸ Gałęzowski D., Sienkiewicz M., *Zarządzanie jakością danych*, Warszawa SGH 06.12.2016 r., <https://studylibpl.com/doc/684763/jako%C5%9B%C4%87-danych>, dostęp 27.07.2019

⁹ Janicki W., *Jakość bazy danych*, w *Automatyka 2005*, Tom 9, Zeszyt 5, Wydawnictwa AGH, Kraków 2005; <http://journals.bg.agh.edu.pl>, dostęp 27.07.2019. Podobne cechy można znaleźć w: Katarzyna Błaszczuk, Knosala Ryszard *Problem jakości danych w hurtowniach*, „Prace Naukowe/Akademia Ekonomiczna w Katowicach”, 2006; <http://www.swo.ae.katowice.pl>, dostęp 27.07.2019.

¹⁰ Tamże

- niedopasowanie dziedziny,
- ewolucja systemów informatycznych,
- konwersje pomiędzy systemami informatycznymi,
- rozszerzenie zastosowania informacji zawartej w bazie danych,
- błędy przetwarzania.

Mieszczą się one w kategoriach defektów wymienionych na Rysunku 1. Oczywiście jest pytanie o możliwość zapewnienia dobrej jakości danych oraz skuteczne wykrywanie defektów. Na ogół zadania te wchodzą w zakres działań działu IT lub wyspecjalizowanej jednostki odpowiedzialnej za jakość danych. Czy jednak mimo wypracowanych rozwiązań systemowych użytkownik – odbiorca danych może być spokojny o ich poprawność, rzetelność? Janicki¹¹ zauważa, że zła jakość danych może mieć wpływ m.in. w obszarach¹²:

- identyfikacji i obsługi klienta,
- przetwarzania danych,
- sporządzania danych sprawozdawczych,
- danych zarządczych,
- działań marketingowych.

W tych obszarach dane służą podejmowaniu rozmaitych decyzji przez użytkownika. Błędne dane prowadzą, na ogół, do błędnych decyzji. Czy zatem decydent jest zdany tylko na przyjmowanie otrzymanych informacji czy też raportów za bezbłędne? Być może istnieje jakiś sposób wykrycia błędów, „defektów danych” w oparciu o samą analizę raportów lub wykresów?

W oparciu o metodę modelowania symulacyjnego oraz autorską technikę „Check Alert” (CA), podjęto próbę wypracowania właśnie takiego rozwiązania systemowego w zakresie wsparcia właściciela danych, w procesie weryfikacji rzetelności analizowanych raportów. Ostatecznie ma ona pozwolić odbiorcy raportów z systemu klasy BI na wskazanie potencjalnych zakresów danych źródłowych, które można podejrzewać, że są błędne.

3. Modelowanie symulacyjne – opis metody badawczej

Modelowanie symulacyjne jest niezwykle użyteczną metodą weryfikacji hipotez definiowanych w ramach wielorakich projektów naukowych, w tym badań realizowanych w

¹¹ Tamże

¹² W praktyce prowadzi to do następujących defektów na poziomie tabel, komórek danych: 1) pominięcie wartości (puste pole); 2) wielokrotne użycie tej samej wartości; 3) pominięcie lub dodanie jednej lub więcej cyfr oraz 4) zamiana kolejności cyfr w ramach tej samej wartości.

zakresie nauk organizacji i zarządzania. Poprzez symulację, modelowanie symulacyjne w niniejszym opracowaniu rozumiemy „naśladowanie, imitowanie rzeczywistego systemu za pomocą eksperymentów przeprowadzonych na modelu reprezentującym (przedstawiającym) ten system”¹³. Niezbędnym elementem umożliwiającym prowadzenie badań, eksperymentów w ramach powyższej metody jest konieczność opracowania, zdefiniowania właściwego modelu – wymiar weryfikacji i walidacji oraz doboru odpowiedniego systemu informatycznego¹⁴.

Istnieje niewątpliwie wiele zalet tej metody, zarówno w wymiarze prowadzonych badań naukowych jak też praktyki menedżerskiej. Należą do nich: duży realizm modelu i danych, możliwość diagnozy rozwiązań/strategii jeszcze nieistniejących, możliwość uwzględnienia parametru czasu/sezonowości, elastyczny dobór kryteriów optymalizacyjnych/ priorytetów/ celów biznesowych, automatyzacja w zakresie testowania wielu różnych wariantów/ scenariuszy biznesowych, minimalizacja wiedzy i umiejętności matematycznych, niski koszt i stosunkowo krótki czas pozyskania wymaganej wiedzy biznesowej. Wadami dla tej metody będą: brak możliwości wypracowania uniwersalnych wniosków i rekomendacji wychodzących poza zakres i przedmiot badań, konieczność przeprowadzenia często wielu eksperymentów – uwarunkowania czasowe, problemy ze wskazaniem rozwiązań optymalnych – kwestia doboru priorytetów, ryzyka związane z manipulowaniem danymi w celu uzyskania oczekiwanych rezultatów czy też weryfikacji wcześniej przyjętych hipotez¹⁵.

W niniejszym opracowaniu, w ramach modelowania symulacyjnego opracowano autorską technikę „Check Alert”, która pozwala przy pomocy systemów klasy BI dokonać oceny jakościowej przetwarzanych danych oraz próbować wskazać źródła występowania tzw. „defektów danych”.

¹³ M. Beaverstock, A. Greenwood, E. Lavery, W. Nordgren, *Symulacja stosowana. Modelowanie i analiza przy wykorzystaniu FlexSim*, FlexSim Software Products, Orem, Utah, USA 2011, s. 6.

¹⁴ „Symulację należy uznać za specyficzną technikę badania rzeczywistości, rozwijaną szczególnie intensywnie w ostatnich dekadach. Rozwój tej techniki był możliwy dzięki rewolucyjnym zmianom w technologii komputerowej (...), dlatego mówiąc współcześnie o symulacji ma się najczęściej na myśli symulację komputerową”. Źródło: Balcerak A., Kwaśnicki W. (2005), *Modelowanie symulacyjne systemów społeczno-gospodarczych: różnorodność podejść i problemów*, [w:] Balcerak A., Kwaśnicki W. (red.) *Symulacja Systemów Społeczno-Gospodarczych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, s. 6-7.

¹⁵ Tamże, s. 6.

4. Zastosowanie techniki „Check Alert” oraz oprogramowania SAP Analytics Cloud w ocenie jakościowej danych biznesowych – studium przypadków

Technika „Check Alert” polega na uważnym przejrzaniu wykresu lub wykresów przedstawiających przetwarzane i analizowane dane oraz zwrócenie uwagi na występujące anomalie. Mogą to być w szczególności:

1. piki, czyli wartości reprezentujące ekstrema globalne, tzn. wartości największe bądź najmniejsze w zbiorze danych leżące wyraźnie poza linią wykresu,
2. wypłaszczenia, czyli kilka występujących obok siebie identycznych wartości,
3. sprzeczność z doświadczeniem, czyli zauważenie, że wygląd wykresu nie potwierdza oczekiwań użytkownika wynikających z doświadczenia i posiadanej wiedzy.
4. dodatkowym sygnałem ostrzegającym użytkownika może być zachowanie wykreślonej linii trendu.

Technika „Check Alert” nie wskazuje w sposób automatyczny na defekt danych, gdyż w rzeczywistości zdarzają się sytuacje niespodziewane powodujące powstanie anomalii na wykresie. Takimi przykładami mogą być: gwałtowne wzrosty lub spadki cen akcji po informacji o zniszczeniu fabryk produkujących komponenty lub kopalni wydobywających surowce strategiczne jako wynik kataklizmu (np. tsunami); bądź informacje o przewrotach wojskowych w krajach o dominującej roli w dostarczaniu jakiegoś surowca powodują powstanie piku na wykresie zmian cen akcji; ataki terrorystyczne na strategiczne obiekty czy też infrastrukturę. Ostateczne rozstrzygnięcie czy obserwowana przez użytkownika anomalia wykresu jest wynikiem defektu danych czy wynika z niespodziewanych zmian rzeczywistości gospodarczej w otoczeniu zewnętrznym lub wewnętrznym musi nastąpić poprzez realizację stosownej procedury z udziałem sekcji jakości danych z działu IT.

W prezentowanych w niniejszym opracowaniu badaniach technikę „Check Alert” wykorzystano do analizy symulacji zdarzeń w trzech różnych firmach (Enea S.A., Orbis S.A. oraz Quantum Software S.A.). Na potrzeby symulacji zmodyfikowano oryginalne dane, tak aby zawierały celowo wprowadzone „defekty danych”¹⁶. Następnie korzystając z

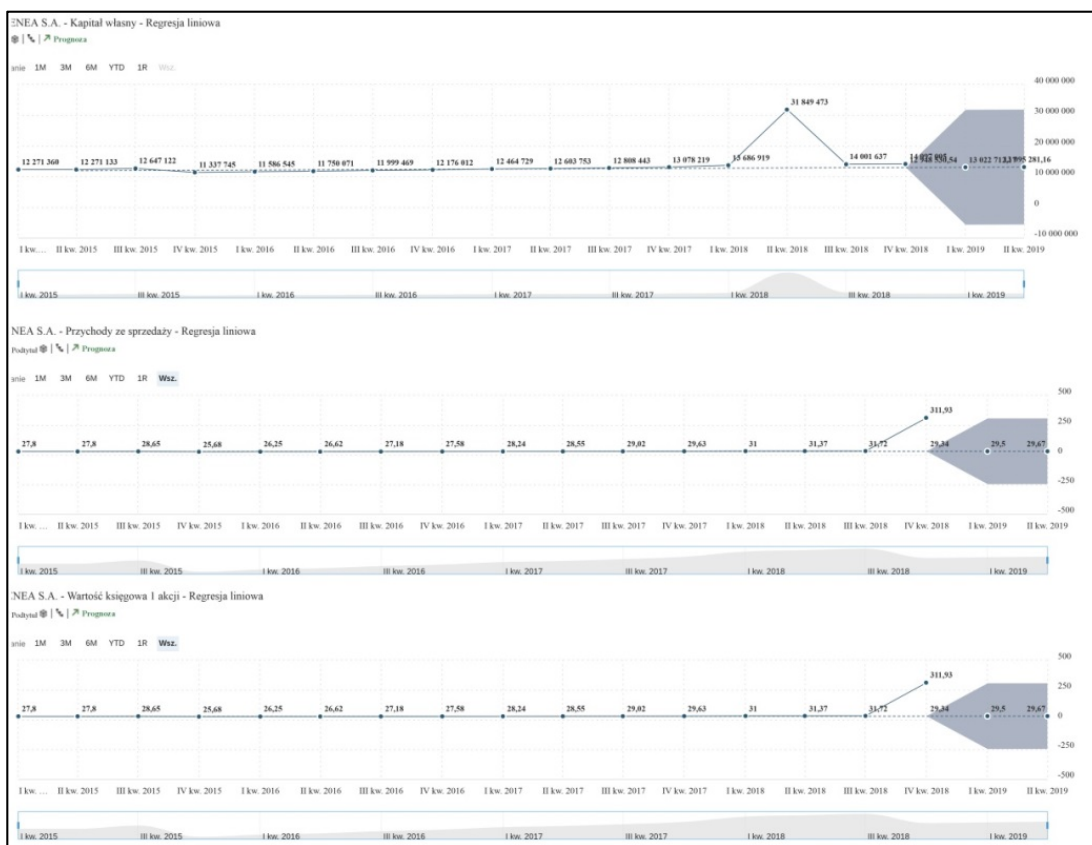
¹⁶ W celu wykonania symulacji dokonano w sposób nieznan dla testera-analityka następujących zmian (defektów) w tabelach danych: 1) Dla spółki Enea S.A., a) dla cechy „Przychody ze sprzedaży” wprowadzono powtórzenie wartości w polu 2016Q2 z 2016Q3; b) dla cechy „Kapitał własny” w polu 2018Q2 wprowadzono zmianę wartości 13 na 31; c) dla cechy „Wartość księgową 1 akcji” w polu 2018Q4 dodano „1” zmieniając wartość z 39,93 na 311,93.2) Dla spółki „Orbis S.A.”, dla cechy „Kapitał własny” dla pola 2018Q4 zmieniono wartość poprzez usunięcie 1 cyfry, tj. z wartości 2 386 425 na 386 425; 3) Dla spółki Enea S.A., dla cechy

zaawansowanego systemu klasy Business Intelligence (*SAP Analytics Cloud*) przygotowano raporty zawierające wykresy z linią trendu i prognozą na dwa kwartały przyszłe¹⁷. W każdym z omawianych przypadków zaprezentowano wykresy, wskazano rodzaj dostrzeżonych anomalii oraz omówiono potencjalne źródło anomalii – źródło defektu.

W punkcie 3 w definicji symulacji mówi się o modelu systemu, natomiast do przeprowadzenia opisanych poniżej eksperymentów wykorzystano wersję komercyjnego systemu udostępnioną przez producenta do celów edukacyjnych oraz rzeczywiste dane zmodyfikowane na potrzeby eksperymentu.

4.1. Firma Enea S.A. – pierwszy zestaw danych

Wykorzystując program SAC¹⁸ wykonano prognozy uwzględniając trzy dostępne metody dla trzech grup danych: kapitał własny, prognoza sprzedaży, wartość księgowa 1 akcji. Kolejne trzy rysunki przedstawiają efekty analizy i prognozy.



¹⁷ „Przychody ze sprzedaży” podczas przygotowania odpowiedniej struktury pliku z danymi wkopiowane dane z cechy „Wartość księgowa 1 akcji” i tym samym dane zostały powtórzone, dwie cechy miały identyczne dane.

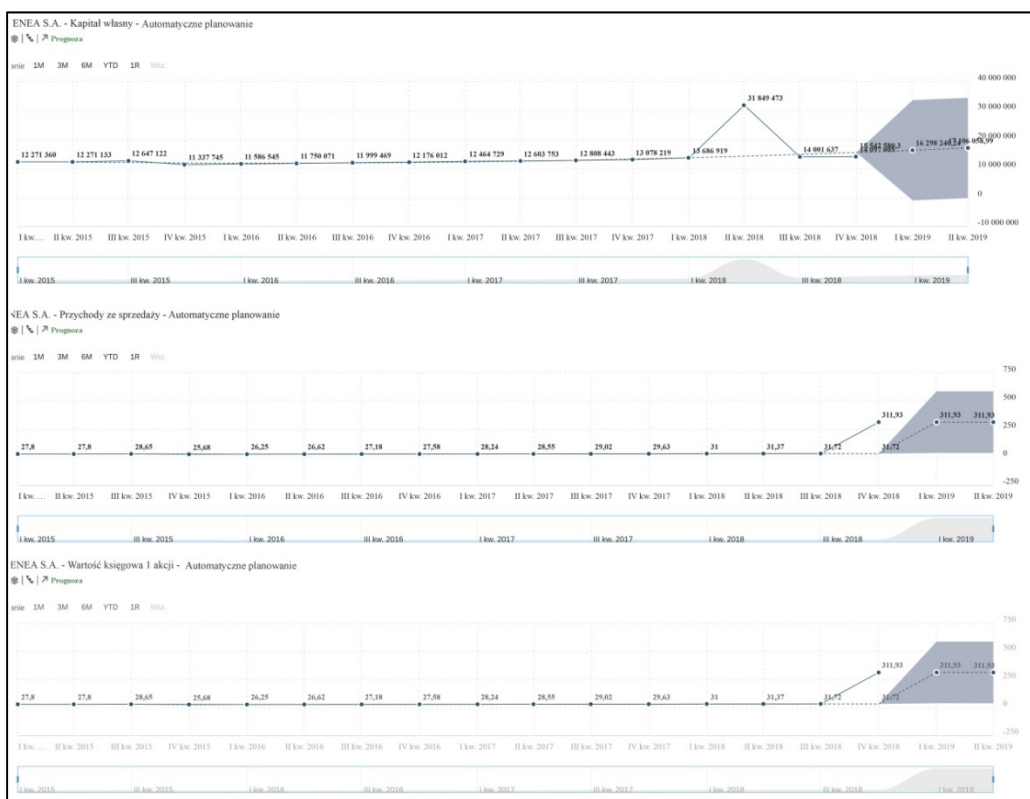
¹⁸ Wykorzystano wszystkie możliwe linie trendu dostępne w systemie w celu zobrazowania możliwości funkcjonalnych programu. Należy jednak pamiętać, że nie ma uniwersalnej metody prognozowania z wykorzystaniem trendu.

¹⁸ Wykorzystano SAC w wersji *SAP Analytics Cloud for Higher Education* z licencją „Academic Account”. Obejmuje ona pełną analizę biznesową, planowanie, analizę predykcijną oraz funkcjonalność *SAP Digital Boardroom* w ramach wspólnego środowiska uczenia się.

Rysunek 2. Enea S.A., metoda: regresja liniowa. Źródło raport systemu SAC.



Rysunek 3. Enea S.A., metoda: potrójne wygładzanie wykładnicze. Źródło raport systemu SAC.



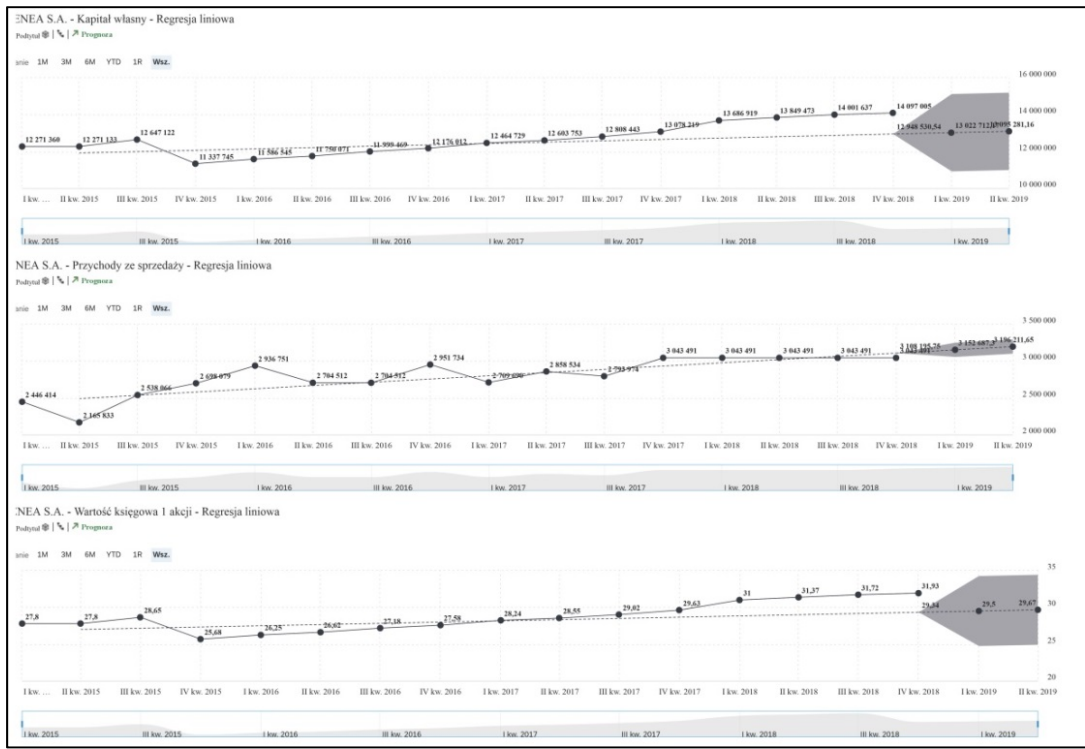
Rysunek 4. Enea S.A., metoda: automatyczne planowanie. Źródło raport systemu SAC.

Odbiorca (użytkownik danych), który analizuje powyższe wykresy powinien zauważyć trzy rzeczy.

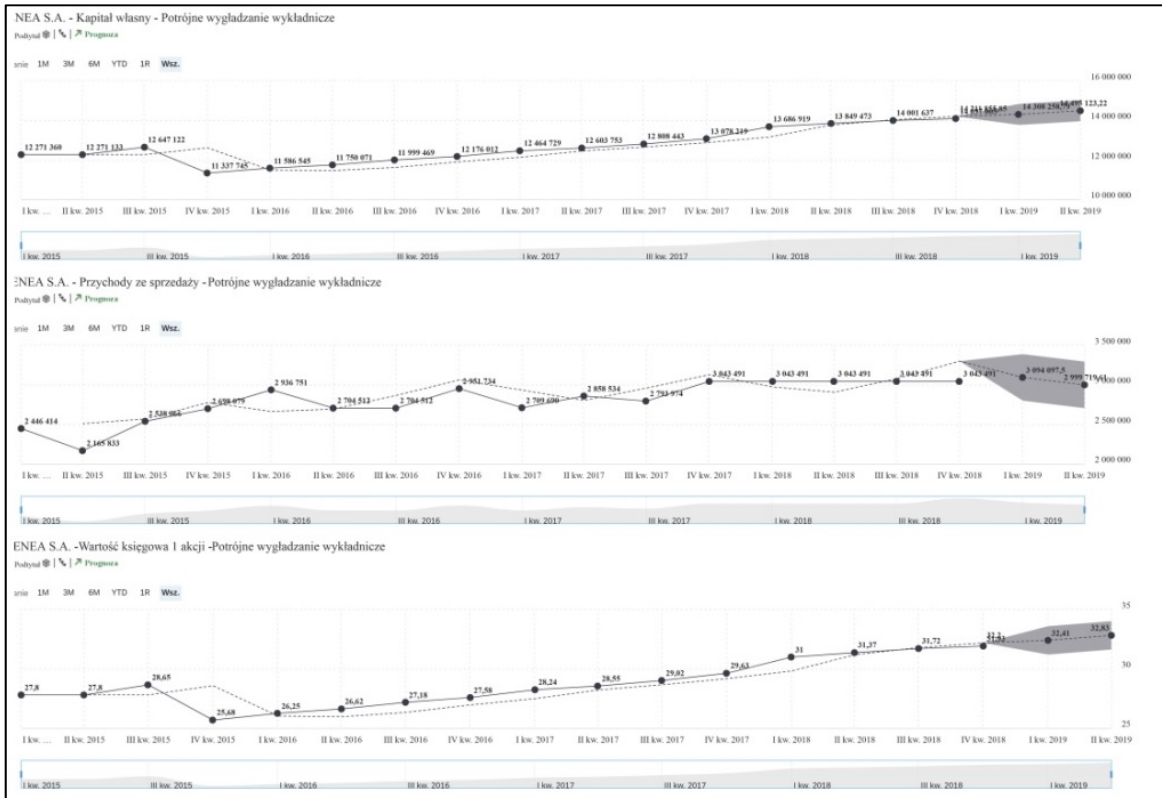
1. Przyglądając się wykresowi kapitału własnego, pierwszy na każdym z rysunków 2-4, widzimy wyraźny skok wartości w II kwartale 2018 r. Linia przerywana, to wykres wartości obliczonych zgodnie z wybraną metodą prognozy, sugeruje znacznie niższą wartość w linii wartości danych źródłowych. Gwałtowny skok kapitału i równie gwałtowny spadek w ciągu kwartału sugeruje błąd danych.
2. Podobnie mamy na wykresie przychodów ze sprzedaży, drugim na każdym z rysunków 2-4, mamy wyraźny skok wartości sprzedaży w IV kwartale 2018 r. To z kolei skutkuje optymistycznym wzrostem sprzedaży w pierwszym półroczu 2019 r. Tu nie jest oczywiste, że skok nie mógł nastąpić, ale jest zbyt duży nawet jak na miesiące jesienno-zimowe. Wartość tego kwartału obliczona zgodnie z wybraną metodą prognozy układa się w linii wartości danych źródłowych, a to sugeruje, że wartość powinna być mniejsza.
3. Wykresy dla przychodów ze sprzedaży i wartości księgowej 1 akcji są identyczne, to oznacza, że dla jednego z wykresów źle przypisano źródło danych. Analizując

wartości danych użytkownik może podejrzewać, że dla przychodów ze sprzedaży przypisano wartości księgowe 1 akcji.

Wszystkie przypadki są efektami potencjalnych defektów danych i powinny być zgłoszone do wyjaśnienia w sekcji zapewnienia jakości danych (do działu IT). Po uzyskaniu potwierdzenia usunięcia defektów użytkownik musi „odświeżyć” wykresy – powtórzyć obliczenia prognozy, zgodnie z obowiązującą procedurą. Odbiorca (użytkownik) otrzymał informację, że defekty danych usunięto i po uruchomieniu raportu uzyskał wykresy jak na rysunkach 5-7.



Rysunek 5. Enea S.A., metoda: regresja liniowa. Źródło raport systemu SAC.



Rysunek 6. Enea S.A., metoda: potrójne wykładnicze. Źródło raport systemu SAC.



Rysunek 5. Enea S.A., metoda: automatyczne planowanie. Źródło raport systemu SAC.

Kolejnym obszarem wymagającym wyjaśnienia jest stała wartość przychodów ze sprzedaży w okresie od IV kwartału 2018 do IV kwartału 2019. Jest mało prawdopodobne aby przez rok przychody ze sprzedaży miały stałą wartość. Jest to zatem kolejny potencjalny błąd będący efektem defektu danych. Odbiorca/użytkownik nie jest w stanie stwierdzić, że występujące na tym samym wykresie dwie identyczne wartości w 2016 r. są potencjalnym wynikiem defektu danych, niekoniecznie tego samego. Wskazane błędy mogą być wynikiem następujących defektów wg klasyfikacji z Rysunku 1:

- błędy użytkowników,
- wady aplikacji,
- migracje wielokrotne,
- starzenie się danych.

Natomiast wg przyczyn wymienionych przez Janickiego¹⁹ mogą to być:

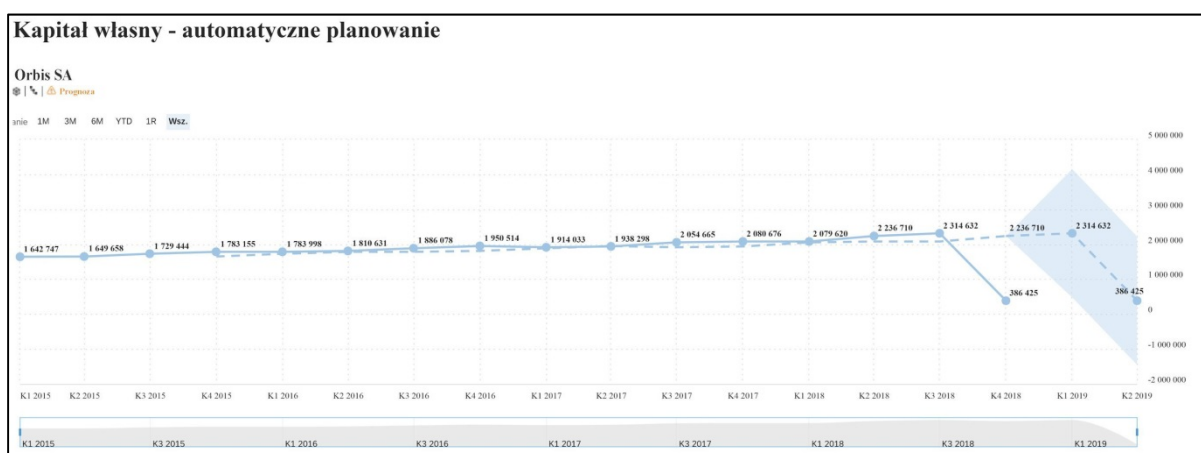
- brak poprawności na etapie wprowadzania,
- ewolucja systemów informatycznych,
- konwersje pomiędzy systemami informatycznymi,
- błędy przetwarzania.

¹⁹ Tamże

Usunięcie defektów i zapobieżenie im w przyszłości leży po stronie sekcji zapewnienia jakości danych (działu IT).

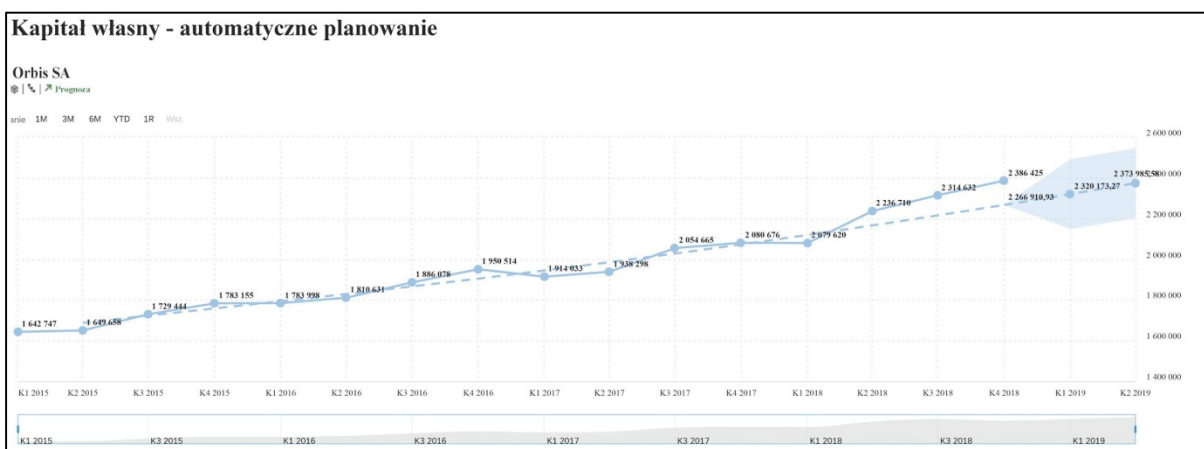
4.2. Firma Orbis S.A. – drugi zestaw danych

Drugi przykład jest analogiczny do defektu rozważanego w podrozdziale 4.1 dla wartości księgowej 1 akcji, dana leży poza przedziałem danych, co na wykresie objawi się pikiem. Korzystając z SAC dokonano analizy danych, tym razem skorzystano tylko z jednej metody obliczania trendu i uzyskano wyniki jak na Rysunku 8.



Rysunek 8. Orbis S.A. – kapitał własny, metoda: automatyczne planowanie. Źródło raport systemu SAC.

Ostatnia dana z IV kwartału 2018 wyraźnie odbiega od linii, w której układają się pozostałe dane. Mamy sytuację analogiczną do wykresy na Rysunku 4, z tym, że dana z ostatniego kwartału 2018 r. leży wyraźnie poniżej linii wykresu. Weryfikacja wartości kapitału własnego z IV kw. 2019 r. potwierdziła błąd i po korekcie otrzymaliśmy wynik jak na Rysunku 9.

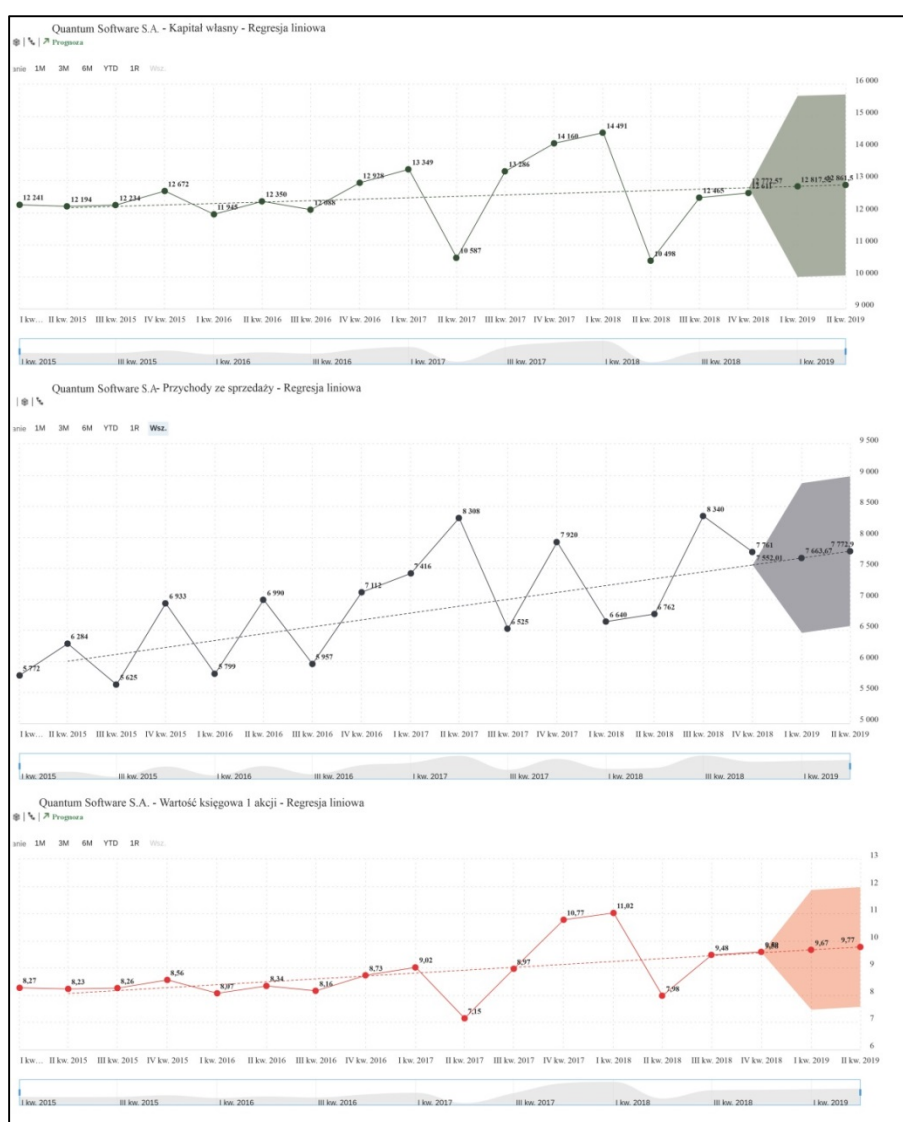


Rysunek 9. Orbis S.A. – kapitał własny, metoda: automatyczne planowanie, dane po korekcie. Źródło raport systemu SAC.

Tak ewidentny przypadek zdarza się rzadko, nie mniej może to być wskazówka weryfikacji danych wprowadzanych „ręcznie”, czyli defekt z kategorii „błędy użytkownika” (Rysunek 1) lub „brak poprawności na etapie wprowadzania” (Janicki²⁰). Przy danych z bazy nie powinno się tak zdarzyć, nie mniej, odczyt do bazy z nieprawidłowo działających urządzeń też może być powodem analogicznego błędu.

4.3. Firma Quantum Software S.A. – trzeci zestaw danych

Rozpatrzmy teraz przykład, który może również zawierać błędne dane, czyli istnienie defektu danych. Ponownie użyliśmy SAC i otrzymaliśmy wynik jak na Rysunku 10.

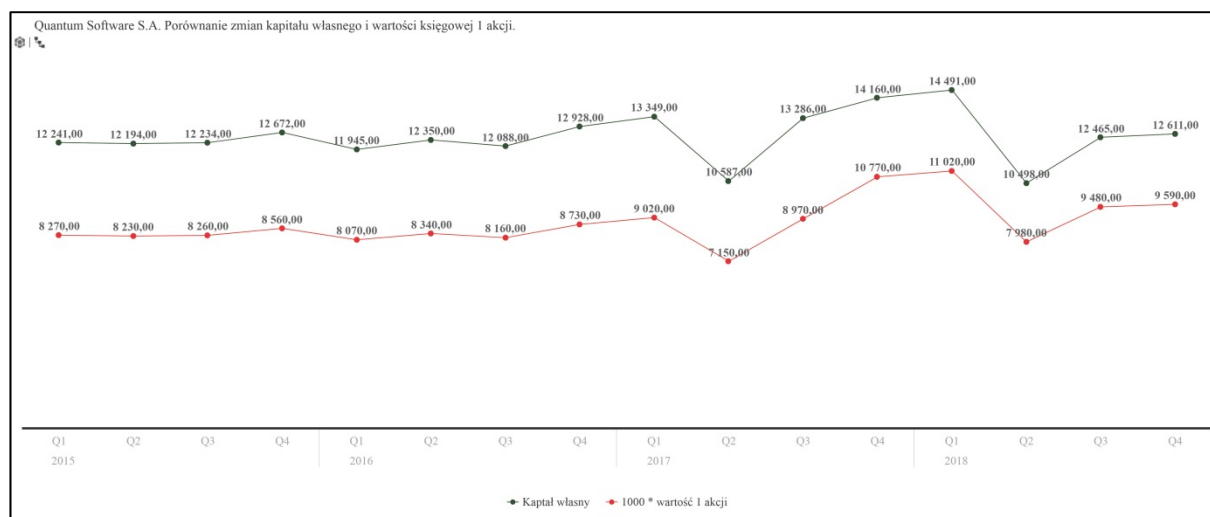


Rysunek10. Quantum Software S. A., metoda: regresja liniowa. Źródło raport systemu SAC.

²⁰ Tamże.

Na wykresach kapitału własnego i wartości księgowej 1 akcji widać piki podobne jak na wykresach kapitału własnego z Rysunków 2-5. Jednak po uważnej analizie zauważymy, że wykresy te są podobne i mają piki w tych samych kwartałach. Łatwo zauważyć korelację wartości księgowej 1 akcji z kapitałem własnym. Zaprezentowano to na Rysunku 11. Natomiast wykres przychodów ze sprzedaży o wielu pikach sugeruje sezonowość sprzedaży.

O tym, że dane nie mają defektu świadczą też stosunkowo małe różnice między kolejnymi kwartałami. W przykładach poprzednich różnice były o rząd wielkości.



Rysunek 11. Quantum Software S. A., wykres porównawczy. Źródło raport systemu SAC.

Ponieważ wartości księgowe 1 akcji są ok. tysiąc razy mniejsze od wartości kapitału własnego, to na wykresie z Rysunku 11 zostały przeskalowane, przemnożone przez 1000, w przeciwnym razie przy wspólnej skali obu wykresów wartości akcji ułożyłyby się w linii prostej. Po operacji skalowania wyraźnie widać podobieństwo wykresów dla rozważanych danych.

Reasumując, po analizie wszystkich opisanych przypadków należy stwierdzić, że zastosowana technika CA nie jest niezawodnym instrumentem wykrywania defektów danych, ale jest bardzo dobrym „papierkiem lakmusowym” na stwierdzenie potencjalnego zagrożenia w zakresie zapewnienia wymaganej ich rzetelności.

Podsumowanie – wnioski i rekomendacje

Na podstawie przeprowadzonej analizy literatury przedmiotu oraz modelowania symulacyjnego, z uwzględnieniem rzeczywistych i zasymulowanych danych biznesowych stwierdzono:

1. O jakości systemu zarządzania informacjami w firmie decyduje zarówno zapewnienie wysokiej sprawności na poziomie produktu informacyjnego jak i samej realizacji procesu [Martyniak].
2. Niezbędnym warunkiem umożliwiającym zapobieganiu pojawianiu się tzw. „choroby informacyjnej” jest wdrożenie koncepcji 8R [Mangan, Lalwani, Butcher] i 5ZRR [Wiktorowska-Jasik].
3. Do najważniejszych przyczyn powstawania defektów danych zaliczamy brak kompletności lub poprawności na etapie ich wprowadzania, brak aktualizacji, niedopasowanie dziedziny, konwersje pomiędzy systemami czy też błędy przetwarzania [Janicki].
4. W wyniku przeprowadzonej symulacji w oparciu o technikę *Check Alert* zauważono, że istnieje realna korzyść w zakresie skutecznego wykrycia potencjalnych defektów danych.
5. Potencjalne wskazania dotyczące anomalii w zakresie wizualizacji danych nie oznacza automatycznie defekt danych i wymaga konsultacji z właścicielem danych.
6. Ocena defektu zależy w dużej mierze od rodzaju potencjalnego defektu, która wymaga indywidualnej diagnozy, np. wartości danych wyznaczone poprzez piki i leżące poza wskazanym przedziałem lub powtarzające się wartości itd.
7. W zakresie masowej obsługi danych biznesowych niezwykle cennym narzędziem jest program klasy Business Intelligence, np. system SAC firmy SAP.
8. Nieznane są inteligentne algorytmy, które w sposób automatyczny mogłyby sygnalizować w formie graficznych alertów (ikon sygnalizacyjnych) defekty danych.
9. Ze sprawnościowej perspektywy zarządzania jakością systemu informacyjnego w firmie jest zdecydowanie lepiej zapewnić wysokie standardy w miejscu powstawania danych (źródła danych) a niżeli potem je wykrywać i poprawiać (zasada „zero defektów”²¹ [Crosby]²²).

²¹ Kluczowa zasada TQM „Do it right the first time”. Źródło: H.H. Steinberg, *Total Quality Management. Kompleksowe Zarządzanie Jakością*, Placet, Warszawa 1998, s. 101.

²² E. Skrzypek, *Jakość i efektywność*, UMCS, Lublin 2000, s. 100.

Reasumując, wykrywanie defektów danych nie jest zadaniem łatwym, tym bardziej, że nie ma skutecznych, automatycznych metod wykrywania anomalii danych pomimo zastosowania systemów informatycznych. Natomiast należy pamiętać, że błędne dane są przyczyną podejmowania błędnych decyzji. Ponieważ w ostatnich latach nastąpiło przesunięcie zakresu obowiązków na stanowiskach pracy z wprowadzania bądź weryfikacji danych w stronę podejmowania decyzji, istotne jest wyposażenie użytkownika/pracownika w techniki wykrywania defektów danych poprzez obserwowane anomalie w ich wizualizacji. Mimo wprowadzania systemów sztucznej inteligencji do podejmowania decyzji, pozostaje wiele obszarów, w których człowiek jest po prostu nie zastąpiony.

Literatura:

Balcerak A., Kwaśnicki W., *Modelowanie symulacyjne systemów społeczno-gospodarczych: różnorodność podejść i problemów*, [w:] Balcerak A., Kwaśnicki W. (red.) *Symulacja Systemów Społeczno-Gospodarczych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2005.

Beaverstock M., Greenwood A., Lavery E., Nordgren W., *Symulacja stosowana. Modelowanie i analiza przy wykorzystaniu FlexSim*, FlexSim Software Products, Orem, Utah, USA 2011.

Mangan, Lalwani, Butcher, *Global Logistics and Supply Chain Management*, Wiley & Sons, 2008.

Martyniak Z., *Podstawy diagnozowania informacji jako produktu i procesu* [w:] Z. Martyniak (red.), *Zarządzanie informacją i komunikacją. Zagadnienia wybrane w świetle studiów i badań empirycznych*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2000.

Skrzypek E., *Jakość i efektywność*, UMCS, Lublin 2000.

Steinberg H.H., *Total Quality Management. Kompleksowe Zarządzanie Jakością*, Placet, Warszawa 1998.

Wiktorowska-Jasik A., *Organizacja procesu transportu ładunków transportem zintegrowanym* [w:] I.N. Semonov (red.), *Zintegrowane łańcuchy transportowe*, Difin, Warszawa 2008.

Beynon-Davies P., *Systemy baz danych*, WNT, Warszawa 1998.

Netografia:

Błaszczyk K., Knosala R., *Problem jakości danych w hurtowniach*, „Prace Naukowe/Akademia Ekonomiczna w Katowicach”, 2006; <http://www.swo.ae.katowice.pl>, dostęp 27.07.2019.

Gałęzowski D., Sienkiewicz M., *Zarządzanie jakością danych*, Warszawa SGH 06.12.2016 r., <https://studylibpl.com>, dostęp 27.07.2019.