



SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
WARSAW SCHOOL OF ECONOMICS

Studium magisterskie

Kierunek: Finanse i Rachunkowość

Specjalność: Finanse przedsiębiorstw

Forma studiów: stacjonarne

Imię i nazwisko: Hubert Guzera

Nr albumu: 61816

**Wykorzystanie modelowania predyktywnego
do stworzenia heurystyki optymalizacji
decyzji podejmowanych w przedsiębiorstwie**

Praca magisterska napisana
w Kolegium Analiz Ekonomicznych
w Katedrze Matematyki i Ekonomii Matematycznej
pod kierunkiem naukowym
dr hab. Michała Ramszy

Warszawa 2015

Spis treści

1	Wprowadzenie	5
2	Cel, założenia i podstawy teoretyczne pracy	7
2.1	Koncepcja pracy	7
2.2	Podstawy teoretyczne	7
2.2.1	Przedsiębiorstwo jako system wieloagentowy	7
2.2.2	Modelowanie predykcyjne	9
2.2.3	Zadanie optymalizacyjne	12
2.3	Proponowany algorytm optymalizacyjny	15
2.4	Cechy algorytmu optymalizacyjnego	16
3	Model	18
3.1	Koncepcja modelu	18
3.2	Zastosowane narzędzia	21
3.3	Struktura programu	21
3.4	Generowanie środowiska modelu	23
3.4.1	Algorytm wyszukiwania drogi	24
3.5	Agenci, ich rodzaje i właściwości	26
3.5.1	Konsumenci	26
3.5.2	Przedsiębiorstwo	28
3.5.3	Produkt	29
3.5.4	Konkurencja	31
3.5.5	Trasy	31
3.6	Symulowanie decyzji konsumenckich	32
4	Efekty działania algorytmu optymalizacyjnego	35
4.1	Założenia modelu	35
4.2	Przewidywanie decyzji konsumentów	39
4.3	Wpływ działania algorytmu na działanie przedsiębiorstwo	40
	Lista tablic	52

Lista rysunków	53
Streszczenie	54

1 Wprowadzenie

Wal-Mart, amerykański gigant handlowy, co godzinę umieszcza w swoich bazach danych 2.5 petabajtów danych, pochodzących z blisko miliona transakcji (The Economist (2010)). I nie jest wyjątkiem — przeciętna ilość danych przechowywanych przez przedsiębiorstwa w Stanach Zjednoczonych jest większa niż zbiory Biblioteki Kongresu (szacowane na 235 terabajtów (McKinsey Global Institute (2011))). W erze informacji większość z naszych działań trafia na serwery tej bądź innej firmy, w formie historii transakcji, koordynatu GPS czy zdjęcia.

Często informacje te zbierane są przypadkiem — ze względu na prowadzenie rachunkowości, specyfikę świadczonych usług, lub też względy archiwizacyjne. Jednak wydobywanie z nich *wiedzy* może stanowić źródło znaczącej przewagi konkurencyjnej. Jak wskazują Brynjolfsson et al. (2011), przedsiębiorstwa podejmujące decyzje na podstawie analizy dużych zbiorów danych (*ang. data driven decision making*) osiągają efektywność o 5-6 proc. większą niż grupa porównawcza. Mają także większy zwrot z kapitału i wycenę rynkową — jednym słowem, radzą sobie lepiej. Nic więc dziwnego, że coraz częściej metody analizy danych *ang. data analytics* stają się priorytetem wśród dużych spółek. Skalę popularności analityki biznesowej *ang. business intelligence* unaocznia badanie PwC (2014), według którego 44 proc. prezesów zarządu planuje oparcie rozwoju firmy o inwestycje w tej dziedzinie.

Ale dzisiejsze zastosowania *big data* to tylko preludium do tego, co czeka nas w przyszłości. Trwający równolegle trend robotyzacyjny spowoduje, że w ciągu 20 lat w przedsiębiorstwie zamiast kierowców możemy zarządzać flotą autonomicznych pojazdów, a magazynierów zastąpią roboty. Fakt, że Google i Daimler już testują takie auta nie pozwala na nazwanie tego science-fiction. Według Carla Freya i Michela Osborne’a z Uniwersytetu w Oxfordzie Frey i Osborne (2013), blisko 47 proc. miejsc pracy jest zagrożonych komputeryzacją. Większość z nich to zawody wykonujące rutynowe, mechaniczne czynności, ale postęp technologiczny powoduje, że na tej liście znajdują się też prace wymagające umiejętności kognitywnych i wnioskowania — jak pracownicy biurowi, analitycy czy operatorzy.

Jeśli więc z jednej strony mamy do czynienia z flotą autonomicznych pojazdów, z drugiej z petabajtami informacji o tym gdzie i co kupują nasi klienci, możemy znaleźć

się w sytuacji, gdzie koordynacja łańcucha dostaw będzie wykraczać poza możliwości człowieka. Dla komputera, wyprognozowanie popytu na podstawie danych i zaplanowanie dostaw nie będzie żadnym problemem. Potwierdza to The McKinsey Global Institute (2011), który wskazuje, że coraz częściej maszyny będą zastępować ludzi w podejmowaniu decyzji i brać udział w sterowaniu przedsiębiorstwem.

W teorii, ze względu na możliwość przeprowadzania złożonych obliczeń i analizy gigabajtów danych, decyzje te będą trafniejsze i poprawią efektywność przedsiębiorstwa.

Niniejsza praca ma na celu skonfrontowanie tej hipotezy. Po pierwsze, poprzez proponowanie jednego z wielu możliwych algorytmów optymalizacji działania przedsiębiorstwa poprzez wykorzystanie istniejących technik modelowania predykcyjnego (*ang. predictive analytics*). Po drugie przez sprawdzenie, jak tak podejmowane decyzje będą wpływać na funkcjonowanie przedsiębiorstwa i czy będzie ono funkcjonować efektywniej, niż gdyby zastosować w nim dotychczasowe praktyki biznesowe.

2 Cel, założenia i podstawy teoretyczne pracy

Praca ma na celu zaproponowanie algorytmu optymalizacji podejmowania decyzji w przedsiębiorstwie na podstawie modelowania predyktywnego oraz sprawdzenie w modelu wieloagentowym, jak zaimplementowanie takiego algorytmu wpływa na wyniki firmy .

2.1 Koncepcja pracy

Rozważane w pracy przedsiębiorstwo należy do sektora szybko zbywalnych towarów konsumpcyjnych (*ang. fast moving consumer goods, FMCG*) i zajmuje się zarówno produkcją, jak i dystrybucją towarów do sklepów detalicznych. Na podstawie historycznych danych o transakcjach i zastosowania modelowania predyktywnego prognozowany będzie krótkoterminowy wolumen sprzedaży w każdym z prowadzonych sklepów. Otrzymana wiedza zostanie wykorzystana do optymalizacji procesów logistycznych, co rozumiane jest przez zbudowanie takich tras dostaw i alokację wśród nich wolumenów produktów, żeby zysk firmy był jak największy.

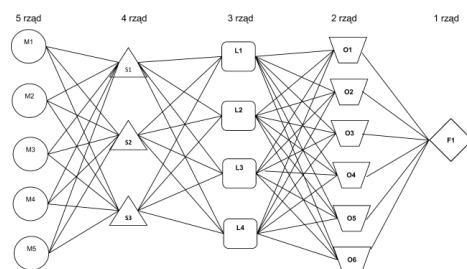
W celu zaprezentowania wyniku działania powstałego w ten sposób algorytmu, zostanie zbudowany model wieloagentowy symulujący rynek i zachowania klientów. Z jego pomocą sprawdzimy wpływ działania algorytmu na decyzje firmy oraz jej wyniki finansowe i operacyjne.

2.2 Podstawy teoretyczne

2.2.1 Przedsiębiorstwo jako system wieloagentowy

Na możliwość wykorzystania modeli wieloagentowych do badania i zarządzania systemami logistycznymi wskazują m.in. Moyaux et al. (2006) czy Kawa (2010). W swoich pracach zauważyli oni, że *producenci*, *dostawcy* i *odbiorcy* i inni uczestnicy łańcucha logistycznego mogą być opisani jako sieć autonomicznych, współpracujących ze sobą agentów, jak pokazane zostało to na wykres 1. Takie podejście, i wynikająca z niego możliwość wykorzystania modeli wieloagentowych pomaga w rozwiązaniu problemów operacyjnych, na jakie wskazuje Kawa (2010). Należy bowiem zwrócić uwagę, że w zakresie wyboru tras i zarządzania flotą wieloetapowe łańcuchy dostaw wielu produktów są problemami NP-trudnymi, szczególnie, że decyzje podejmowane lokalnie są współza-

leżne.¹ Ponadto, jak zauważa Kawa (2010), w sieci przedsiębiorstw pomiędzy dostawcami kolejnych rzędów (tj. fabryki, magazyny, sklepy) może istnieć wiele połączeń które są wobec siebie konkurencyjne, ponieważ jeden magazyn może zaopatrywać się w wielu fabrykach. Zastosowanie w tej dziedzinie modeli wieloagentowych pozwala więc na zbadanie, jak decyzje podejmowane na jednym z etapów łańcucha dostaw wpłyną na cały system i innych uczestników.



Wykres 1: Przedsiębiorstwo przedstawione jako graf kierunkowy. Źródło: Kawa (2010)

W niniejszej pracy stosowane jest rozszerzenie tego podejścia, poprzez zaprogramowanie jako agentów jednostek organizacyjnych przedsiębiorstwa (*fabryka, magazyn, sklep, zarząd*), które razem tworzą system (*przedsiębiorstwo*).

To podejście opiera się na obserwacji, że relacje pomiędzy jednostkami w przedsiębiorstwie są analogiczne do relacji uczestników łańcucha dostaw. Porter (1985) zauważył, że działalność przedsiębiorstwa to de facto sekwencja działań, która na każdym ogniwie zwiększa wartość dla odbiorcy. Zasady funkcjonowania opisywanego przez Portera *łańcucha wartości* są identyczne co do opisywanego przez Moyaux et al. (2006) i Kawa (2010) łańcucha dostaw a relacje pomiędzy nimi można przedstawić w sposób zaproponowany przez Kawa (2010) — z wykorzystaniem grafu skierowanego, jak zaprezentowano na wykresie 1. Podobieństwo to podkreśla fakt, że przedsiębiorstwa poprzez strategię *integracji wertykalnej* swym zasięgiem mogą w rzadkich przypadkach objąć całość łańcucha dostaw.

¹Decyzje podjęte na krańcowych etapach łańcucha wpływają na wcześniejsze lub późniejsze etapy, co Moyaux et al. (2006) opisuje jako m.in. "bullwhip effect".

Ponieważ przedsiębiorstwa często dysponują wieloma duplikującymi swoje działania jednostkami ², łańcuch ten jest nieliniowy i w jego przypadku mamy do czynienia z podobnymi wyzwaniem co w łańcuchu logistycznym.

Zdefiniowanie jako agentów poszczególnych jednostek przedsiębiorstwa jest przy tym spójne z określoną przez Wooldridge i Jennings (1995) charakterystyką agenta, który według ich postulatów posiada :

- autonomię — poszczególne jednostki przedsiębiorstwa podążają za strategią i celami narzuconymi przez zarząd, ale mają zazwyczaj swobodę w podejmowaniu decyzji mających na celu ich realizację,
- zdolności do komunikacji — jednostki przedsiębiorstwa komunikują się z otoczeniem (relacje z klientami) oraz między sobą (raportowanie do zarządu, spotkania), a w ramach pomiędzy jednostkami przedsiębiorstwa istnieje asymetria informacji,
- reaktywność — jednostki przedsiębiorstwa reagują na zmiany rynkowe oraz zmiany wewnątrz przedsiębiorstwa,
- proaktywność — jednostki przedsiębiorstwa podejmują inicjatywy mające na celu zwiększyć wartość przedsiębiorstwa, jak działalność innowacyjna bądź ekspansja.

Dlatego w niniejszej pracy będziemy rozważać model wieloagentowy, w którym według założeń na przedsiębiorstwo składać się będzie szereg autonomicznych agentów (*fabryka, magazyn, sklep, zarząd*), wspólnie tworzących łańcuch dostaw.

2.2.2 Modelowanie predykcyjne

W procesach logistycznych i produkcyjnych kluczowym wyzwaniem jest niepewność związana ze zmiennością sprzedaży i jej wartości w chwili $t+1$. Jak wskazuje James et al. (2013) do zmniejszenia niepewności poprzez prognozowanie sprzedaży możemy wykorzystać metody statystyczne. Według James et al. (2013), zakładając, że dysponujemy zbiorem n obserwacji p zmiennych, możemy zbadać ich relację ze zmienną wyjaśnianą y i otrzymać *model*, który dla nowych - nieanalizowanych wcześniej - obserwacji $x_1, x_2...x_n$

²Dobrym przykładem są tutaj zakłady samochodowe, które mogą produkować dany model w różnych krajach. Zmiana fabryki powoduje przy tym radykalną zmianę łańcucha dostaw

zwraca przewidywaną wartość zmiennej objaśnianej \hat{y} . Różnorodne metody modelowania zmiennej objaśnianej (wspólnie nazywane przez James et al. (2013) uczeniem statystycznym, *ang. statistical learning*) mogą być wykorzystane do modelowania predykcyjnego, tj. przewidywania przyszłych wydarzeń na podstawie przeszłej historii danych, w tym również przyszłego wolumenu sprzedaży w przedsiębiorstwie.

Zastosowanie metod *statistical learning* w przedsiębiorstwach potwierdza Buckinx et al. (2007), który wskazywał na możliwość prognozowania lojalności klienta na podstawie wewnętrznych danych o transakcjach, oraz Davenport et al. (2011), który opisuje szereg *case studies* firm, w których wykorzystuje się istniejące dane o transakcjach do przewidywania przyszłych zakupów klientów. Jednym z podanych przez niego przykładów jest Tesco, które na podstawie zebranych danych przewiduje, jak będzie wyglądał koszyk zakupów klienta podczas następnych zakupów, i odpowiednio wcześniej wysyła mu bony zniżkowe. Również podczas panelu *Strategia B2C w erze Big Data - jak wykorzystać potencjał danych* na XXV Forum Ekonomicznego w Krynicy przedstawiciele polskiego biznesu zwracali uwagę na szerokie wykorzystanie modelowania predykcyjnego również w nasz gospodarce.³

Podane przykłady dają nam podstawy, żeby w przypadku optymalizowanego przedsiębiorstwa zakładać, że dane o każdej transakcji są zapisywane wraz z niektórymi danymi osobowymi klienta⁴, a cały zbiór danych może być wykorzystany do przewidywania sprzedaży w chwili $t + 1$.

Dlatego w pracy będziemy zakładać, że dla każdej transakcji w sklepie dysponujemy zbiorem informacji, zawierające dane o transakcji (*data, miejsce, rodzaj płatności*), produkcie (*nazwa produktu, cena, ilość*) oraz dane osobowe klienta (*płeć, wiek, zarobki, wykształcenie*).

³Jednak jak zwracano uwagę, *big data* i *modelowanie predykcyjne* służą głównie do zyskiwania wiedzy o kliencie i rynku do manualnego przetworzenia analiz, a nie automatyzacji podejmowania decyzji, co rozważamy w tej pracy

⁴To stwierdzenie opiera się na opisywanym przez Davenport et al. (2011) przypadku Tesco i zastosowanej przez nich metody zbierania danych. Możliwość zbierania danych o transakcjach określonego klienta daje karta lojalnościowa (lub konto, w przypadku e-commerce), na którą rejestrowana jest każda transakcja. Praktyką jest, żeby przy okazji tworzenia karty lojalnościowej zbierać informacje o kliencie w ankiecie (zakres danych zależy od praktyki korporacyjnej). Dzięki temu, przedsiębiorstwa są w stanie przypisać do zakupów dane osobowe jak płeć, miejsce zamieszkania, wykształcenie etc.

Na podstawie takiego zbioru danych chcemy przewidzieć popyt na produkty w każdym ze sklepów, co w praktyce oznacza konieczność stworzenia modelu estymującego następujące wielkości

- liczebność poszczególnych grup klientów ⁵ odwiedzających sklep w chwili $t + 1$,

oraz, w zależności od zastosowanego podejścia,

- prawdopodobieństwo z jakim klient o danej charakterystyce kupi produkt,
- produkt wybrany przez danego klienta.

Jak wskazuje James et al. (2013), zmienna objaśniana może przyjąć różne dziedziny — m.in. zmiennej binarnej (1/0), prawdopodobieństwa, *log odds* lub klasy — i z tego względu każdy z przypadków różni się metodami które możemy zastosować. Według sugestii James et al. (2013) i Hastie et al. (2001), zastosujemy następujące metody

- do liczby klientów - regresję metodą OLS (*ordinary least squares, metoda najmniejszych kwadratów*)
- do prawdopodobieństwa zakupu - regresję logistyczną (*logistic regression*)
- do wyboru produktu- metody klasyfikacyjne *k-means* oraz *drzewa klasyfikacyjne*

oraz metody nie służące bezpośrednio do modelowania \hat{y} , jednak wspierające proces predykcyjny

- pomiar odległości (*ang. distance scaling*), poprzez liczenie odległości euklidesowej (*euclidean distance*) pomiędzy dwoma zbiorami danych i stworzenie obliczenie macierzy niepodobieństwa (odległości, *dissimilarity matrix*)
- eliminację zmiennych (*ang. backward elimination*), która jest jednym z podejść wyboru podzbiorów (*subset selection*) do selekcji zmiennych wyjaśniających które wspólnie tworzą najlepszy model
- prawdopodobieństwo warunkowe do przewidywania, jacy konsumenci odwiedzają sklep w $t + 1$.

⁵Przez *poszczególne grupy klientów* rozumiemy klientów o wspólnej charakterystyce, czyli takich samych zestawach zmiennych identyfikujących $\langle kl_1..kl_n \rangle$

2.2.3 Zadanie optymalizacyjne

Rozważmy przedsiębiorstwo, które za argumentacją przedstawioną w rozdziale 2.2 oraz obserwacjami Moyaux et al. (2006) oraz Kawa (2010) przedstawiamy jako system składający się z niezależnych agentów. Zakładamy, że rozważane przedsiębiorstwo składa się z

Definicja przedsiębiorstwa i grafu

- zbioru *fabryk* oznaczony przez FA , z generycznym elementem $fa_n \in FA$
- zbioru *magazynów* oznaczony przez MA , z generycznym elementem $ma_n \in MA$
- zbioru *sklepów* oznaczony przez SK , z generycznym elementem $sk_n \in SK$
- *zarządu*, pełniący rolę centralnego koordynatora

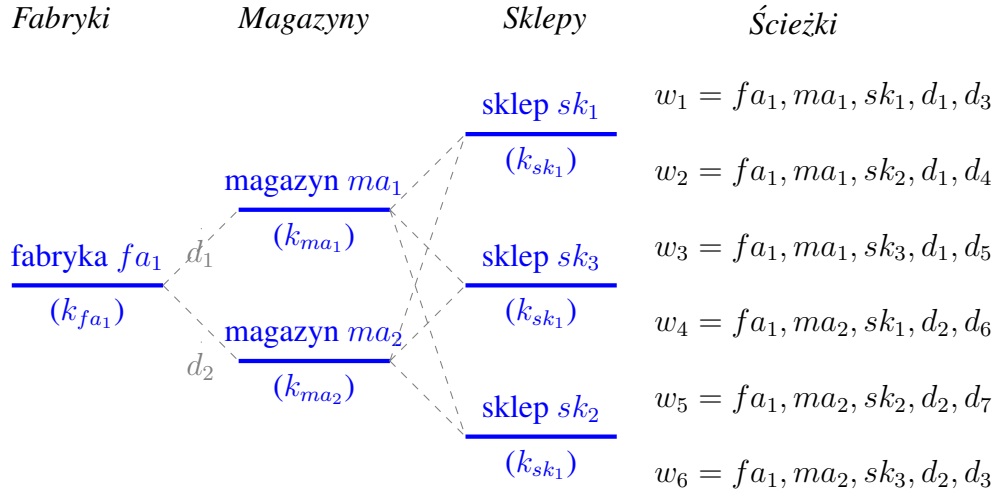
Jak zauważył Kawa (2010), jeśli rozpatrujemy przedsiębiorstwo pod kątem procesów logistycznych możemy zaobserwować, że jednostki przedsiębiorstwa będą wspólnie tworzyć graf kierunkowy $S = \{FA \cup MA \cup SK \cup D\}$ (zob. wykres 2), gdzie jednostki przedsiębiorstwa FA, MA, SK będą wierzchołkami grafu, a trasy dostaw pomiędzy jednostkami będą krawędziami grafu $d \in D = \{(i, j) : i \in FA, j \in MA\} \cup \{(i, j) : i \in MA, j \in SK\}$.

Planowanie produkcyjno-logistyczne w przedsiębiorstwie będzie polegało na alokacji wolumenów produkcji na poszczególne ścieżki $w \in W = (fa, ma, sk, d_1, d_2)$ ⁶.

Należy zaznaczyć, że optymalizacja alokacji ścieżek zamiast krawędzi ma znaczenie praktyczne dla przedsiębiorstw, szczególnie międzynarodowych. Dzisiejsze trendy globalizacyjne spowodowały, że często produkcja w danym kraju trafia na wiele rynków, a każda z partii musi być dostosowana do rynków lokalnych⁷ — czyli w momencie produkcji produkt musi mieć określone miejsce docelowej dostawy. W logice grafu oznacza to,

⁶Ścieżka zawiera dwa elementy ze zbioru krawędzi d , ponieważ zawiera krawędź (fabryka — magazyn) oraz (magazyn — sklep)

⁷Ze względów zarówno regulacyjny, jak i preferencji klientów. Dodatkowo, w niektórych branżach, jak samochodowej, można dostosowywać specyfikację zamawianego produktu. W takim przypadku również należy zapewnić, że trafi on do miejsca docelowego



Wykres 2: Rozważane przedsiębiorstwo przedstawione jako graf kierunkowy.

Źródło: opracowanie własne

że przedsiębiorstwo musi określić ruch na całej ścieżce produktu, nie tylko do następnej krawędzi.

Równanie zysku przedsiębiorstwo

W tak zdefiniowanym grafie proces logistyczny odbywa się przepływem wzdłuż krawędzi D , a z każdym z elementów grafu S związany jest koszt przepływu $k_j = f_j(x_j)$ ⁸, gdzie f jest dowolną funkcją kosztu, j rozważanym elementem grafu S , a x_j wolumenem produkcji przechodzącym przez dany element.

Sprzedaż towarów przepływających przez graf następuje w sklepach, a zakładając, że przedsiębiorstwo nie stosuje dyskryminacji cenowej, cena produktu będzie globalna i stała⁹. Przychód r w każdym ze sklepów $sk \in SK$ przy cenie p będzie równy $r = p \times q_{sk}$, gdzie q to sprzedaż w wybranym sklepie sk ¹⁰.

⁸Warto zauważyć, że koszt przesyłu nie oznacza tylko kosztów transportu, ale także kosztów produkcji w fabryce, kosztów magazynowania oraz kosztów obsługi procesów sprzedaży w sklepie, tak więc koszt ten będzie definiowany nie tylko na krawędziach D , ale każdym elemencie grafu S

⁹Rozumiemy przez to, że cena jest identyczna dla wszystkich klientów, dla wszystkich sklepów oraz wszystkich okresów czasu t

¹⁰Warto zauważyć, że sprzedaż q_j nie jest tym samym co wolumen n_j , ponieważ dostarczenie towaru do sklepu nie gwarantuje jego sprzedaży

Dla całego systemu funkcja zysku systemu P_s będzie zależała przede wszystkim od alokacji wolumenów zadany równaniem 1

$$P = \sum_{sk} p \times q_{sk} - \sum_j f_j(x_j) \quad (1)$$

gdzie p to stała cena,

q_{sk} to sprzedaż w sklepie $sk \in SK$ spełniająca warunek $q_{sk} \leq x_{sk}$

a f_j to dowolna funkcja kosztu zależna od wolumenu x_j w elemencie grafu $j \in S$

To zadanie optymalizacyjne byłoby względnie proste ¹¹, gdybyśmy mieli doskonałą informację na temat poziomu sprzedaży w każdym ze sklepów w chwili $t + 1$. Na taką wiedzę nie możemy liczyć ani w tej pracy, ani w rzeczywistości, dlatego rozwiązaniem proponowanym w niniejszej pracy jest zastosowanie modelowania predykcyjnego (*predictive analytics*), w celu prognozowania liczby klientów i ich wyborów w każdym ze sklepów w najbliższych okresach czasu. ¹²

Ponadto, jak warto zauważyć, nie możemy liczyć na to, że funkcja kosztu f_j będzie liniowa. Wspominana przez Kawa (2010) maksymalna przepustowość każdego z łańcuchów, która w przedsiębiorstwie będzie spowodowana ograniczonymi mocami produkcyjnymi, może spowodować, że funkcja kosztu będzie dowolną funkcją, co empirycznie potwierdzone jest w powszechności efektów skali w każdym z sektorów gospodarki. Warto przy tym zauważyć, że efekty skali mogą być zarówno dodatnie, jak i ujemne, jak również ich wspólne występowanie ¹³.

Trzecim aspektem który trzeba wziąć pod uwagę jest złożoność obliczeniowa. Nawet dla prostego układu, lecz wolumenu produkcji ponad 1000 sztuk sprawdzenie zysku w przypadku wszystkich kombinacji alokacji wymaga olbrzymiej ilości iteracji. Mimo znaczącego wzrostu mocy komputerów w ostatnich latach, wolumeny produkcji w dużych przedsiębiorstwach oraz złożoność tras logistycznych sprawia, że to podejście jest kompletnie niepraktyczne i należy szukać alternatywnych podejść, upraszczających problem.

¹¹Pomijając aspekt złożoności obliczeniowej

¹²Obecnie w przedsiębiorstwach rzadko stosuje się zaawansowane sposoby prognozowania sprzedaży (*predictive analytics*), a zarządzanie dostawami odbywa się raczej metodą manualnego uzupełniania zapasów.

¹³Dodatnie gdy produkcja będzie mniejsza niż optymalny poziom produkcji, a ujemne po przekroczeniu optymalnych mocy produkcyjnych

Proponowany algorytm będzie brał wszystkie te aspekty pod uwagę, a pierwszy z wymienionych problemów rozwiążemy wykorzystując *modelowanie predykcyjne*.

2.3 Proponowany algorytm optymalizacyjny

W proponowanym algorytmie optymalizacyjnym wykorzystujemy fakt, że dzięki modelowaniu predyktywnemu możemy prognozować wolumen sprzedaży q_{sk} w każdym ze sklepów $sk \in SK$. Ponieważ równolegle zakładamy brak stosowania dyskryminacji cenowej, część przychodowa równania 1 będzie nam znana, i możemy ją oznaczyć jako stałą r_s , oznaczającą przychód całego systemu.

$$P_s = r_s - \sum_j f_j(x_j) \quad (2)$$

Ponieważ w rozdziale 2.2.3 stwierdziliśmy, że ze względów biznesowych naszym celem jest optymalizacja alokacji wolumenu na ścieżkach, zauważmy, że x_j będzie równe $\sum_{w:j \in w} x_w$, sumie wolumenów x_w na wszystkich ścieżkach $w \in W$ przechodzących przez element grafu $j \in S$, dzięki czemu równanie możemy zapisać jako:

$$P_s = r_s - \sum_j f_j\left(\sum_{w:j \in w} x_w\right) \quad (3)$$

Dzięki temu przekształceniu możemy wyznaczyć gradient funkcji zysku...

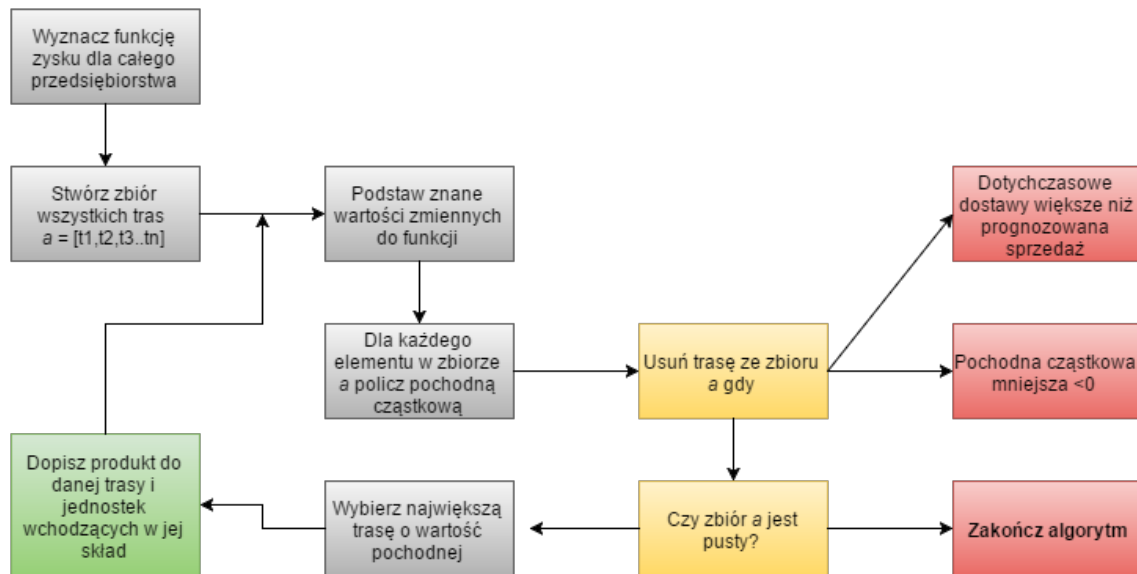
$$\nabla P_s = \left[\frac{\partial P_s}{\partial x_{w_1}} \dots \frac{\partial P_s}{\partial x_{w_n}} \right] \quad (4)$$

gdzie P_s to funkcja zysku, x_w to wolumen przechodzący przez ścieżkę $w \in W$,

a n to ilość ścieżek w zbiorze W

... który wskaże nam kierunek najszybszych wzrostów wartości funkcji zysku P_s w zależności od ulokowania kolejnego produktu na którejś ze ścieżek. Zakładając, że działanie algorytmu zaczyna się, gdy każda ze ścieżek $w \in W$ jest pusta, tj. każde $x_w = 0$, działanie algorytmu jest następujące

1. wyznacz gradient funkcji zysku P_s
2. stwórz wektor A wszystkich ścieżek $w \in W$, który będzie nam służył do iterowania pętli



Wykres 3: Proponowany algorytm optymalizacyjny

3. dla każdego produktu lub partii produktu oblicz gradient w punkcie $[x_{w_1}, x_{w_n}]$.
4. dla każdego elementu wektora A sprawdź,
 - czy odpowiadająca mu wartość pochodnej cząstkowej jest mniejsza od 0
 - czy dodanie kolejnego produktu do ścieżki spowoduje, że w sklepie leżącym na ścieżce przestanie być spełniany warunek $q_{sk} \leq x_{sk}$ ¹⁴
5. jeśli którykolwiek z powyższych jest prawdą, usuń ścieżkę z wektora A
6. jeśli wektor A nie jest pusty, z pozostałych elementów wektora A , wybierz ścieżkę o największej wartości pochodnej cząstkowej w gradiencie
7. powtórz od kroku 3. dopóki wektor A nie będzie pusty.

Działanie algorytmu w standardzie UML zaprezentowane zostało na wykresie ??.

2.4 Cechy algorytmu optymalizacyjnego

Algorytm działający według reguł opisanych w rozdziale 2.3 będzie heurystyką, ponieważ kierując się najszybszymi wzrostami w punkcie obliczania gradientu może ominąć maksimum globalne. Gwarantuje za to znalezienie maksimum lokalnego i zakończenie pętli

¹⁴Wolumen dostaw będzie większy niż prognozowany wolumen sprzedaży

¹⁵. Biorąc pod uwagę, że w aspekcie praktycznym większość funkcji kosztów nie będzie miała nietypowych wzorów, algorytm też powinien być w zupełności wystarczający do większości zastosowań optymalizacyjnych.

Ponieważ algorytm nie oblicza wszystkich możliwych kombinacji alokacji wolumenów, a używa gradientu do oszacowania najbardziej opłacalnej trasy dla każdego z produktów, czas jego wykonania będzie znacznie krótszy w stosunku do algorytmu kompletnego. Liczba iteracji potrzebnych do kalkulacji zysku w przypadku każdej możliwej kombinacji alokacji będzie liczbą Stirlinga II rodzaju, a w przypadku algorytmu maksymalna liczba iteracji będzie iloczynem liczby produktów oraz liczby ścieżek.¹⁶

Warto zauważyć, że w sytuacji, gdy skala działalności firmy powoduje, że graf jest bardzo rozbudowany, sposób działania algorytmu pozwala predefiniować w macierzy A najbardziej prawdopodobne trasy. Pozwoli to zmniejszyć liczbę iteracji, jednocześnie zachowując integralność grafu - inne metody wymagałyby zmiany jego struktury.

Innym ograniczeniem algorytmu jest to, że szukając kolejnych optymalnych punktów "porusza się" dyskretnie po wykresie funkcji. Wynika to z założenia, że w procesach logistycznych nie możemy przekroić i przetransportować pół produktu, a nawet ładunek drobnicowy będzie miał swoją jednostkę najmniejszego możliwego transportu, jak paleta, tona, kontener etc.

¹⁵Tj. nie ma możliwości, żeby iterował w nieskończoność

¹⁶W praktyce będzie mniejsza, ponieważ algorytm pozbywa się nierentownych i zapełnionych ścieżek

3 Model

W celu sprawdzenia działania algorytmu zbudowany został model wieloagentowy, który symuluje lokalny rynek na wybrany produkt, wraz z zachowaniami konsumentów i funkcjonowaniem przedsiębiorstwa.

3.1 Koncepcja modelu

Zgodnie z argumentacją zawartą w rozdziale 2.2, relacje w przedsiębiorstwie pomiędzy jednostkami tworzącymi wspólnie łańcuch dostaw można przedstawić jako model wieloagentowy. Jednocześnie, inspirując się Kaminski (2012), zakładamy, że możemy stworzyć heterogenicznych agentów oraz symulować ich decyzje celu modelowania zachowań i trendów na rynku.

Łącząc te podejścia, celem pracy jest stworzenie modelu wieloagentowe symulującego rynek (w szczególności kładąc nacisk na aspekt sprzedaży oraz dostaw) na którym moglibyśmy sprawdzić wpływ działania algorytmu na wyniki przedsiębiorstwa.

Zgodnie z powyższym, w modelu znajdują się następujące typy agentów:

- **klienci**, którzy zgodnie z założeniem będą heterogeniczni i definiowani przez cechy demograficzne ¹⁷, które wpływają na podejmowane przez niego decyzje oraz zachowania,
- **przedsiębiorstwo**, sprzedające *produkt* na rynku. Zgodnie z podejściem przedstawionym w 2.2, przedsiębiorstwo będzie rozumiane jako zbiór niezależnych, współpracujących ze sobą agentów
 - fabryk
 - magazynów
 - sklepów
 - oraz zarządu, pełniącego funkcje koordynującą

¹⁷Są to między innymi wiek, zarobki, wykształcenie, zainteresowania - zostanie to dokładnie opisane w dalszej części pracy.

- **konkurencji**, zachowującej się pasywnie w stosunku do rynku, konsumentów i symulowanego przedsiębiorstwa, ale wprowadzającej na rynek szereg produktów stanowiących alternatywę dla produktu symulowanej firmy ¹⁸
- **produktów** dostępnych na rynku, z których każdy zdefiniowany jest unikalnymi cechami określającymi jakość, typ i cenę produktu, przez co każdy z produktów będzie preferowany przez inną grupę konsumentów, a preferencje są oparte na danych ze świata rzeczywistego.

Aspekt lokalizacji

Aby dobrze odwzorować kluczowy aspekt lokalizacji i dróg w łańcuchach dostaw, symulowany rynek jest osadzony w *wirtualnym mieście*. Oznacza to, że każdy agent ma swoją lokalizację w macierzy o wymiarach $x \times y$ i może się w niej poruszać po wyznaczonych drogach.

Lokalizacja wpływa na działania agenta - klient kupi produkt tylko w sklepie który będzie na jego ścieżce, a dostawa z magazynu do sklepu będzie tym droższa, im bardziej oddalone będą od siebie.

Symulowanie decyzji konsumenckich

W modelu konsumenci nieustannie poruszają się po *mapie*, bez związku z działaniem przedsiębiorstwa ¹⁹. W każdej jednostce czasu t klienci z prawdopodobieństwem p będą potrzebować symulowany produkt. Wywołanie tego zdarzenia spowoduje, że podczas losowej podróży z punktu A do B odwiedzą oni losowy sklep z wszystkich sąsiadujących z trasą, a następnie wybiorą jeden z produktów dostępnych w sklepie ²⁰.

Symulacja wyboru opiera się na danych o preferencjach konsumenckich zebranych w grze ekonomicznej na próbie 169 badanych, w wyniku których otrzymano 1860 rekordów

¹⁸Tj. konkurencja nie zmienia decyzji podjętych przed rozpoczęciem gry, i w założeniu ma stanowić jedynie alternatywę dla konsumentów.

¹⁹Ruchy są wywołane przez losowe zdarzenia którym może być poddany konsument. Zdarzenia wymagają od niego podróży do jednego z predefiniowanych miejsc — jak praca czy dom innego agenta

²⁰Sklepy nie przynależą do przedsiębiorstwa, więc znajdują się tam także produkty konkurencji, i spośród wszystkich klient dokonuje wyboru.

danych ²¹. Na ich podstawie których zbudowane zostało drzewo klasyfikacyjne opisujące prawdopodobieństwo zakupu produktu od cech konsumenta.

Ponieważ każdy konsument-agent w modelu ma swoje unikalne cechy demograficzne i charakteru, spójne z danymi zebranymi w ankiecie, wykorzystujemy zbudowane drzewo klasyfikacyjnego do określenia wyboru, jakiego najprawdopodobniej w świecie rzeczywistym dokonał by jego odpowiednik, posiadający identyczne bądź zbliżone cechy. ²² Przeprowadzając podobny proces dla każdego konsumenta w modelu, otrzymujemy dynamiczną symulację rynku produktów szybkozbywalnych.

Decyzje przedsiębiorstwa

Ponieważ, jak zostało wspomniane, konsument wybiera produkt tylko z gamy dostępnych w sklepie, kluczowe dla sukcesu przedsiębiorstwa w modelu jest dostarczenie w każdej jednostce czasu t odpowiedniej ilości produktów ²³. Oznacza to, że przed rozpoczęciem każdej tury przedsiębiorstwo musi podjąć szereg decyzji o m.in.

- odpowiednim poziomie produkcji,
- wolumenie dostaw do każdego ze sklepów w sieci,
- rozdzieleniu wolumenów pomiędzy jednostki przedsiębiorstwa, tj. ile z całkowitego wolumenu ma wyprodukować fabryka A , a ile fabryka B
- ścieżki jakimi powinny zostać wysłane dostawy do każdego ze sklepów. ²⁴

Każda z tych decyzji będzie miała wpływ na przychody ²⁵ oraz koszty firmy. W modelu przedsiębiorstwo przez n jednostek czasu t przedsiębiorstwo korzysta z predefinio-

²¹Gra ekonomiczna została dokładniej opisana w rozdziale 3.6

²²Oczywiście, o wiele lepsze byłoby oparcie pracy o prawdziwe historie transakcji, jednak jest to niemożliwe ze względu na dużą poufność tych danych

²³W warunkach symulacji nie można zapełnić *półek sklepowych* do pełna, bo brak sprzedaży oznacza stratę dla przedsiębiorstwa, a w sektorze FMCG zakładamy, że w czasie $t + 1$ produkty tracą zdolność do spożycia

²⁴Rozumiemy przez to pytanie, który z magazynów ma być pośrednikiem, ponieważ dobra nie mogą być dostarczane bezpośrednio z fabryki do sklepu.

²⁵Przy założeniu stałej ceny, będą to przede wszystkim *utraczone koszty* w przypadku wyczerpania się zapasów w sklepie

wanych zasad podejmowania decyzji, które oparte są na najczęstszych praktykach spotykanych w świecie rzeczywistym:

- towar zamawiamy jest zawsze z najbliższego magazynu
- ilość zamówionego towaru do każdego ze sklepów równa jest sprzedaży t_0 ²⁶

Po n rund, przez pozostałą ilość okresów t decyzje w przedsiębiorstwie podejmowane są na podstawie algorytmu optymalizującego, opisanego w rozdziale ??.

3.2 Zastosowane narzędzia

Model został zbudowany w języku programowania Python 2.7, z wykorzystaniem następujących bibliotek (zob.tabela 1) :

Tabela 1: Zastosowane biblioteki języka Python. *Źródło:* opracowanie własne.

Biblioteka	Źródło	Zastosowanie
Sympy	www.sympy.org	Wykorzystanie do obliczeń symbolicznych
scikit-learn	scikit-learn.org	Wykorzystanie bibliotek metod statystycznych

Kod programu dostępny jest pod adresem github.com/hubertguzera/master-thesis

3.3 Struktura programu

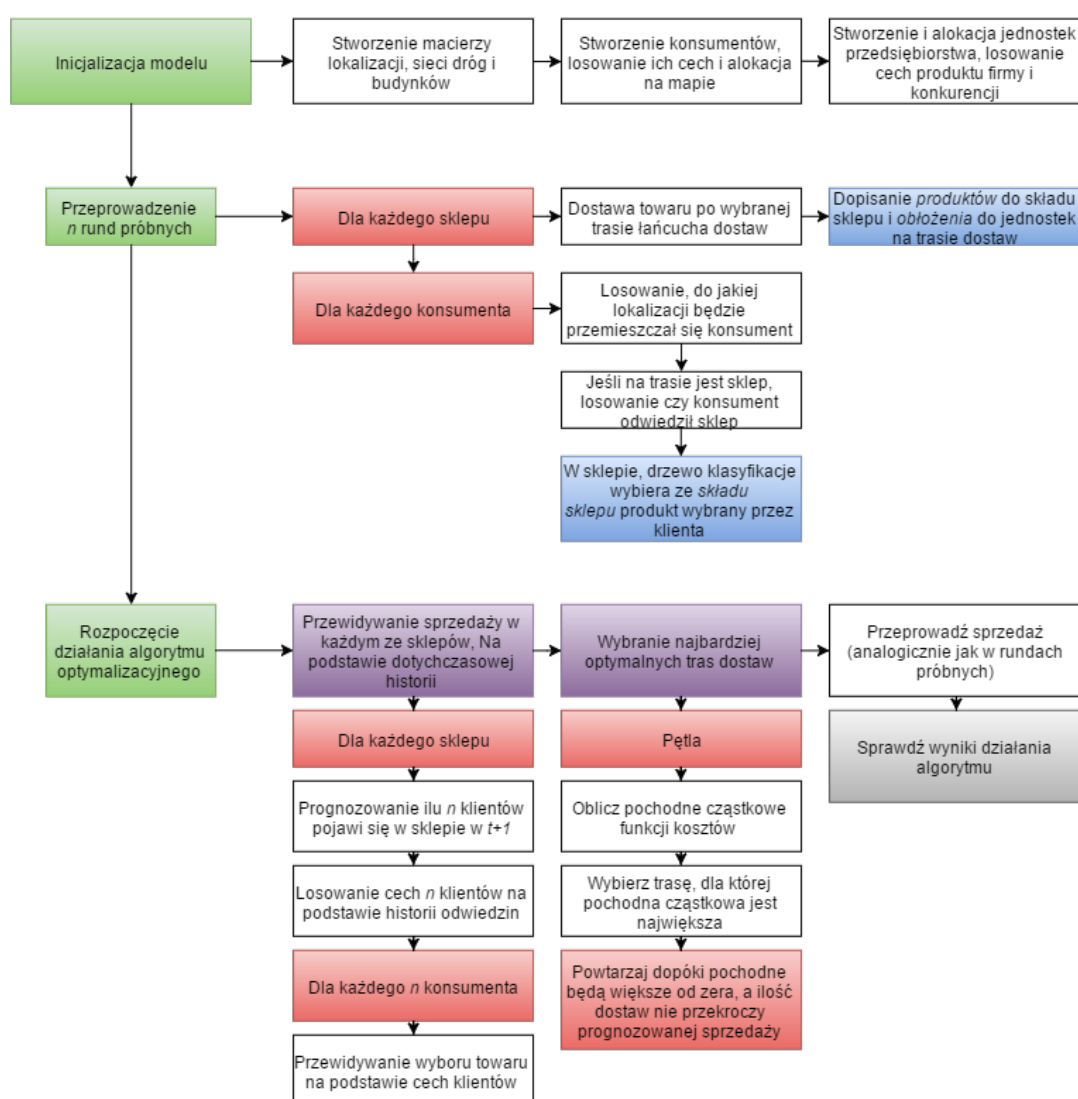
Program podzielony jest na trzy moduły, jak zaprezentowano na rysunku ??

- Pierwsza odpowiada za stworzenie, w drodze losowań, środowiska w ramach którego toczy się symulacja, wraz z agentami i macierzą lokalizacji. ²⁷

²⁶Gdzie t_0 to runda próbna, która ma na celu sprawdzenie popytu na rynku i nie jest zapisywana do wyników firmy. Jej wdrożenie ma na celu odzwierciedlenie w modelu wiedzy powszechnej o rynku, tj. przedsiębiorcy wiedzą ile mogą sprzedać na podstawie tego, co sprzedawali w niedalekiej przeszłości albo na podstawie raportów rynkowych

²⁷Model może pominąć tą część i wczytać pregenerowany świat w celu sprawdzenia różnych scenariuszy w statycznym świecie (*ceteris paribus*).

- Druga część przez t_n jednostek czasu symuluje działanie rynku — decyzji klientów i funkcjonowania przedsiębiorstwa — według predefiniowanych zasad sterowań²⁸, zwracając na koniec wyniki (przychody, koszty i zysk) przedsiębiorstwa w danej jednostce czasu t .
- Trzecia po t_n rund symulacji aplikuje algorytm optymalizacyjny, który poprzez prognozowanie sprzedaży na krańcach grafu w t_{n+1} , szuka najlepszych decyzji o alokacji produktów. Działanie algorytmu można porównać do miary porównawczej uzyskanej w pkt 2.



Wykres 4: Struktura działania programu Źródło: opracowanie własne

²⁸Co istotne, algorytm optymalizacyjny jeszcze nie jest implementowany na tym etapie

3.4 Generowanie środowiska modelu

Klasa *rynek*

Za reprezentację tak opisanego środowiska modelu odpowiada klasa *rynek*, wobec której dziedziczą wszystkie inne klasy występujące w modelu. Klasa *rynek* (i wszystkie dziedziczące) jest generowana dynamicznie i losowo ²⁹. Konstrukcja klasy *rynek* w programie została zaprezentowana w diagramie 5.

rynek
swiat : class
symulowana_firma : class
tura : integer
produkty_na_rynku : class
__init__(self,swiat) : None
sprzedaz_w_sklepach(self) : None
nowatura (self) : None

Wykres 5: Diagram UML klasy *rynek* Źródło: opracowanie własne

Klasa *świat*

Klasa *swiat* zawarta w klasie *rynek* i zaprezentowana na diagramie 10 powstała w celu odpowiedniego odwzorowania kluczowego aspektu lokalizacji i dróg w łańcuchach dostaw, wzorując się na podejściu zastosowanym w *modelu segregacji Schellinga* (Schelling (1971)). Agenci osadzeni są w przestrzeni, reprezentowaną przez macierz klas *lokalizacja* o wymiarach (x,y). Dodatkowo, lokalizacje są połączone drogami, wymuszając na agentach poruszanie się tylko w obrębie ścieżek. Dzięki temu, w modelu będziemy mogli wiernie odwzorować wpływ odległości i wyboru trasy na efektywność procesów logistycznych, oraz zależność wyników sklepu od zamieszkującej okolicę populacji.

²⁹Jednak może być zapisana jeśli istnieje konieczność replikacji obliczeń albo porównań.

Macierz *mapa* generowana jest generowana jest według algorytmu widocznego na rysunku 7, który gwarantuje następujące właściwości otrzymanej w ten sposób macierzy lokalizacji.

- Drogi krzyżują się i skręcają tylko pod kątem prostym. Poza skrzyżowaniami, drogi nie mają w sąsiedztwie innych dróg,
 - Inne lokalizacje (domy, sklepy etc.) mogą występować tylko w bezpośrednim sąsiedztwie drogi,
 - Drogi stanowią ciągłą linię, dzięki czemu nie ma punktu, do którego nie dałoby się dojechać z dowolnego miejsca startowego,
 - W regionie 2 punktów od skraju mapy nie są generowane ani drogi, ani lokalizacje.
- 30
- Gęstość sieci dróg oraz prawdopodobieństwo występowania zakrętów jest predefiniowana przez użytkownika

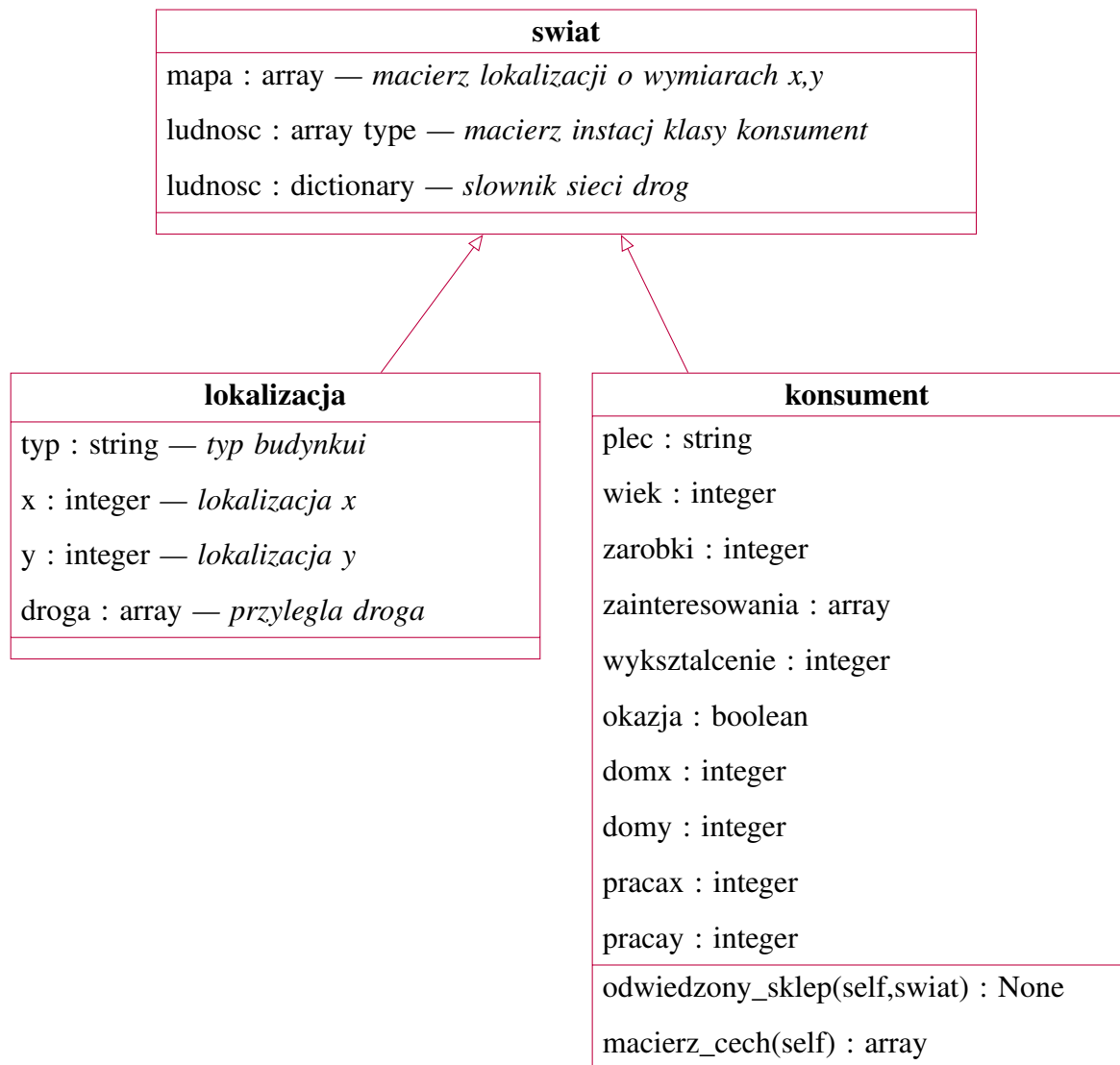
Przykładowa *mapa* otrzymana w wyniku działania algorytmu widoczna jest na rysunku 8.

3.4.1 Algorytm wyszukiwania drogi

Trasy pomiędzy zadanymi punktami w modelu są wyszukiwane dynamicznie, na podstawie algorytmu wyszukiwania drogi. Odległości z kolei obliczane są poprzez sumowanie ilości punktów w zwracanym przez algorytm łańcuchu.

Algorytm oparty jest na metodach wyszukiwania ścieżek w grafach, dzięki założeniu, że każda droga o współrzędnej (x,y) na mapie jest punktem grafu, który może sąsiadować

³⁰Jest to zabezpieczenie algorytmu, który w odległości 2 pkt od skraju mapy ma 0 proc. szansy na poprowadzenie ścieżki dalej - ponieważ w przypadku iterowania na skrajach mapy niektóre funkcje (jak sprawdzenie sąsiadujących punktów) mogą odnieść się do lokalizacji poza macierzą, powodując błąd programu



Wykres 6: Diagram UML klasy swiat, lokalizacja i konsument *Źródło:* opracowanie własne

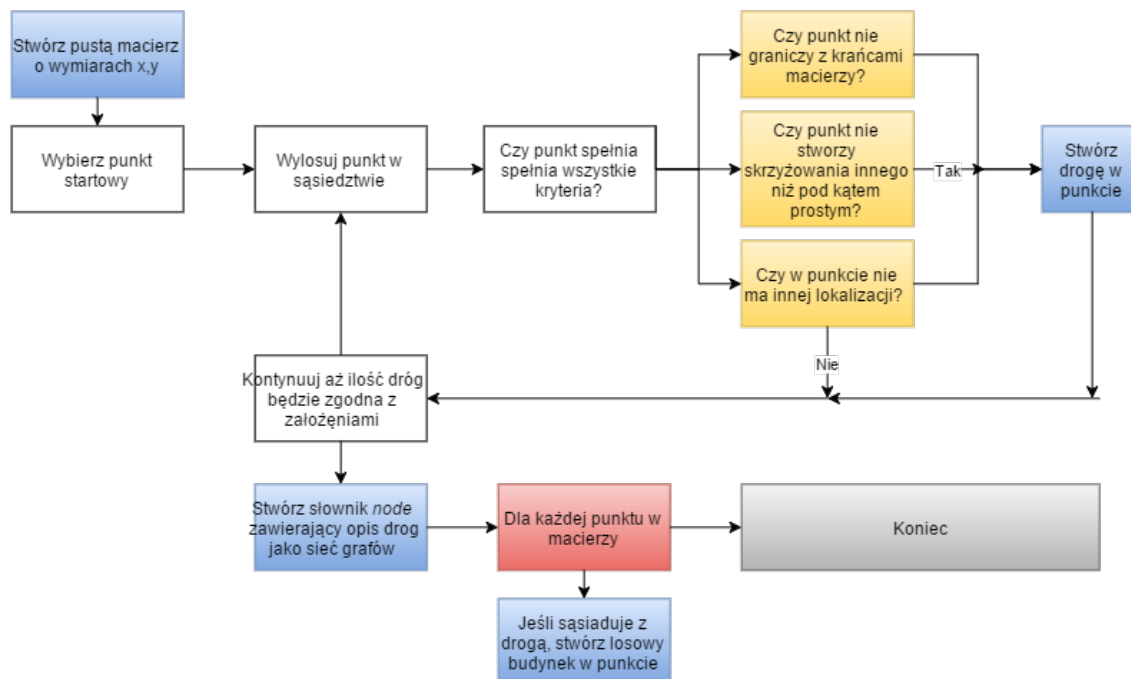
z punktami o współrzędnych $(x-1,y),(x+1,y),(x,y+1),(x,y-1)$ ³¹, o ile również są drogami.

32 33

³¹Punkty $(x-1,y-1),(x+1,y-1),(x+1,y+1),(x-1,y+1)$ wykluczamy przez wcześniejsze założenie, że drogi krzyżują się tylko pod kątem prostym

³²Informacje o punktach i sąsiadujących przechowywane są w zmiennej nodes, która jest słownikiem, dla każdego klucza - punktu na mapie - przechowuje informacje o sąsiadujących punktach, np. $(3,2) = [(3,3)(4,3)]$.

³³Pewnym ograniczeniem jest, że jako punkty grafu definiujemy tylko drogi, tak więc szukając trasy z punktu A do punktu B de facto szukamy trasy z drogi przy punkcie A do drogi przy punkcie B.



Wykres 7: Algorytm generowania świata w modelu. *Źródło:* opracowanie własne

Algorytm, udostępniony przez Python Foundation ³⁴ i zaprezentowany w tabeli ?? ma następujące cechy

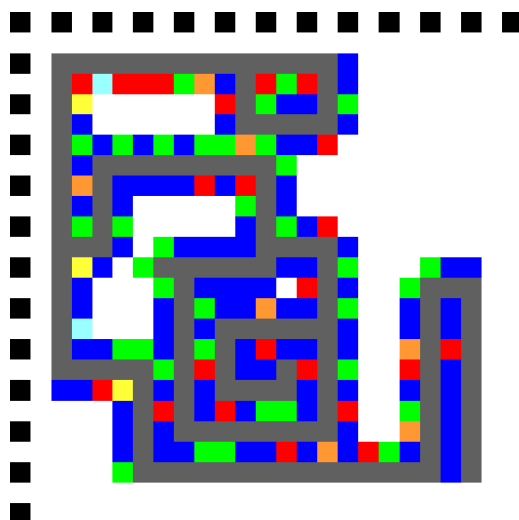
- jest rekurencyjny,
- nie jest losowy,
- nie gwarantuje znalezienia najkrótszej trasy.

3.5 Agenci, ich rodzaje i właściwości

3.5.1 Konsumenci

Idąc za Kaminski (2012), w modelu stosujemy modelowanie rynku za pomocą heterogeniczne konsumentów. Stąd, każdy z konsumentów ma swoją unikalną charakterystykę rozumianą przez cechy demograficzne oraz cechy charakteru, które będą wpływać na jego wybory.

³⁴<https://www.python.org/doc/essays/graphs/>



Wykres 8: Przykładowa mapa - *szary* to drogi, *niebieski* - domy mieszkalne, *czerwony i zielony* - biurowce, *pomarańczowy* - sklepy, *żółty* - magazyny, *błękitny* - fabryki *Źródło:* opracowanie własne

W dodatku, chociaż agenci nieustannie poruszają się w ramach modelu, to pula lokalizacji w ramach których będą się przemieszczać jest ograniczona. Każdy z konsumentów będzie się przemieszczał tylko w wyniku predefiniowanych zdarzeń, których liczba dla każdego konsumenta jest ograniczona do czterech ³⁵, tak więc agent będzie przemieszczał się pomiędzy maksymalnie czterema lokalizacjami, tworząc pewne wzorce zachowań.

Dzięki temu, odzwierciedlamy zjawisko ze świata rzeczywistego, że konsumenci zazwyczaj robią zakupy w ograniczonej liczbie sklepów będących po drodze bądź niedaleko. Jest to bardzo istotny warunek funkcjonowania modelu, ponieważ losowy dobór klientów uniemożliwiłby modelowanie predykcyjne.

Każdy będzie definiowany w klasie o następujących właściwościach

Wartości dla każdego z konsumentów są losowane niezależnie na podstawie rozkładów publikowanych przez Główny Urząd Statystyczny Główny Urząd Statystyczny (2011) oraz danych firmy Sedlak&Sedlak (Sedlak&Sedlak (2013)) ³⁶ w celu zagwarantowania odzwierciedlenia struktury społeczeństwa. Ze względu na zastosowanie prawdopo-

³⁵Warto zauważyć, że każde ze zdarzeń wywołuje przemieszczenie do korespondującej lokalizacji, tak zdarzenie będzie zawsze wywoływać podróżę tej samej lokacji

³⁶Raporty firmy Sedlak&Sedlak służyły do zbudowania tabeli prawdopodobieństwa wystąpienia danego wynagrodzenia w zależności od płci i wykształcenia. Reszta danych oparta na GUS

```

1  def find_path(graph, start, end, path=[]):
2      path = path + [start]
3      if start == end:
4          return path
5      if not graph.has_key(start):
6          return None
7      for node in graph[start]:
8          if node not in path:
9              newpath = find_path(graph, node, end, path)
10             if newpath: return newpath
11  return None

```

Wykres 9: Algorytm wyszukiwania drogi Źródło: opracowanie własne

dobieństw warunkowych dla niektórych cech (np. zarobki są zależne od wykształcenia) istnieje pomiędzy nimi korelacja.³⁷ Przykładowe rozkłady pokazane są na rysunku 11a oraz 11

Poza zaprezentowanymi na histogramach danymi demograficznymi agenci posiadają trzy zainteresowania, których rozkład zaprezentowany został w tabeli ???. W przeciwieństwie do danych demograficznych, są one niezależne, a każda z nich posiadała identyczną szansę na wylosowanie. Celem tego było wprowadzenie do modelu zmiennych, których nie byłyby skorelowane z innymi, a ponieważ mają wpływ na wybory klientów i nie są zapisywane w historii transakcji, utrudniają przewidywanie sprzedaży.

3.5.2 Przedsiębiorstwo

Zgodnie z założeniami określonymi w rozdziale 1, symulowane przedsiębiorstwo jest w istocie zbiorem niezależnych agentów. Zbiór ten należy w modelu do klasy *firma*, przechowującej w macierzach instancje klas *fabryka*, *magazyn*, *sklep*, jak zaprezentowano na diagramie ???. W klasie *firma* znajdują się także funkcje przynależne opisywanemu w strukturze modelu zarządowi, mające na celu koordynację działania przedsiębiorstwa.

Warto zauważyć, że klasy te mają wiele wspólnych właściwości, z wyjątkami obecnymi w klasie *sklep*. Wynika to z konieczności stworzenia funkcji symulujących sprzedaż

³⁷Chociaż korelacja może być problemem przy modelowaniu, będziemy sobie z nią radzić na późniejszym etapie

konsument
<p>plec : string</p> <p>wiek : integer</p> <p>zarobki : integer</p> <p>zainteresowania : array</p> <p>wykształcenie : integer</p> <p>okazja : boolean — <i>wskazanie powodu wyjścia z domu</i></p> <p>domx : integer — <i>współrzędna x domu</i></p> <p>domy : integer — <i>współrzędna y domu</i></p> <p>pracax : integer — <i>współrzędna x pracy</i></p> <p>pracay : integer — <i>współrzędna y pracy</i></p>
<p>odwiedzony_sklep(self,swiat) : None</p> <p>macierz_cech(self) : array</p>

Wykres 10: Diagram UML klasy konsument *Źródło:* opracowanie własne

oraz faktu, że sklepy mają dodatkowe zmienne przechowujące informacje o składzie towaru, klientach odwiedzających sklep w danej jednostce czasu t oraz historii transakcji.

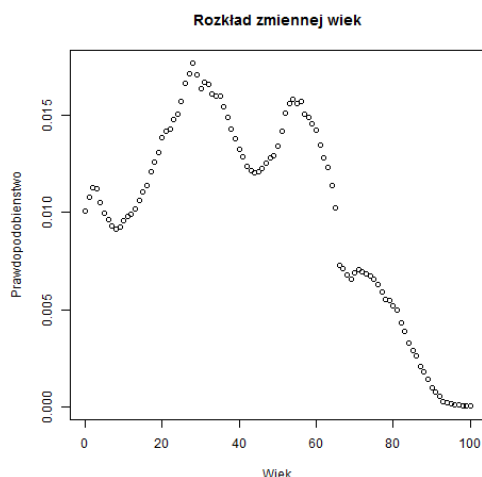
3.5.3 Produkt

Opierając się na stwierdzeniu Sagan (2011), który wskazuje na nurtu w dziedzinie modelowania strukturalnego zachowań klientów, które wykorzystywało zmienne marketingowe określające jakość produktu i ilość cech ³⁸, zakładamy, że każdy produkt charakteryzuje się cechami wpływające na prawdopodobieństwo jego zakupu przez konsumentów które mogą być wyrażone ilościowo.

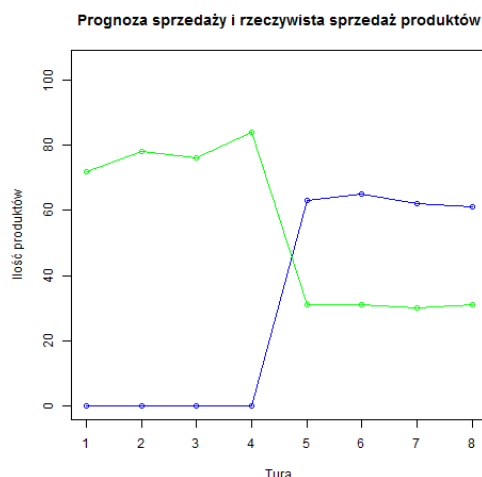
Stąd produkt definiujemy przez zestaw wybranych cech które odróżniają go od produktów konkurencji, które mogą przyjąć formę skali ocen, zmiennych binarnych bądź zmiennych kategoriycznych. ³⁹, a klasa *produkt* przechowujące jednowymiarową macierz z cechami produktu.

³⁸Mowa o tzw. nurcie poznawczym albo nurcie teorii przetwarzania informacji - TPI

³⁹Ich istotność nie jest w tym momencie ważna, ponieważ nawet jeśli w zbiorze znajdzie się cecha mająca mały wpływ na decyzje konsumentów, zostanie ona wyeliminowana na etapie tworzenia modelu bądź drzewa klasyfikacyjnego ze względu na brak istotności statystycznej współczynnika



(a) Rozkład prawdopodobieństwa zmiennej wiek



(b) Rozkład prawdopodobieństwa zmiennej wykształcenie

Wykres 11: Rozkłady zmiennych cech klasy konsument *Źródło:* opracowanie własne na podstawie danych Narodowego Spisu Ludności 2011, Główny Urząd Statystyczny (2011)

Warto odnotować, że w pracy przyjmujemy, że symulowanym produktem z branży FMCG jest piwo. Ten dość nieelegancki wybór motywowany jest głównie specyfiką produktu, która dobrze pasuje do wymagań stawianych przez model (szczególnie w aspekcie możliwości modelowania decyzji konsumenckich), wśród których są

- wysoka sprzedaż i idący za tym wysoki obrót towaru w sklepach — przeciętny Polak pije 99 litrów piwa rocznie, co oznacza butelkę kupioną co mniej więcej drugi dzień,⁴⁰
- produkt mają silne marki o ugruntowanych cecach i grupach docelowych — reklamy piw zazwyczaj podkreślają w jakich okolicznościach i jaka grupa docelowa pije piwo, co powoduje, że występuje wysoka zależność pomiędzy charakterystyką demograficzną klienta a piwem które wybierze,⁴¹.

⁴⁰Dzięki temu możemy założyć, że prawdopodobieństwo zakupu piwa przez klienta to nawet 50% , a to z kolei gwarantuje odpowiednią ilość iteracji do przeprowadzenia analizy.

⁴¹Idéalnym przykładem jest Redd's, wybierany głównie przez kobiety. Innymi mogą być Grolsch wybierany przez ludzi zamożnych, kiedy Wojak trafia do najgorzej zarabiających

- produkty mocno różnią się od siebie cechami — w przeciwieństwie do mąki, rola jakości, smaku i właściwości piwa odgrywa znaczącą rolę dla klienta. Powoduje to, że piwo smakowe i pszeniczne będą trafiać do dwóch różnych grup klientów.

Tabela 2: Cechy charakteryzujące produkt w modelu. *Źródło:* opracowanie własne.

Cecha	Rodzaj zmiennej	Skala	Opis
Cena	Liczba naturalna	1-5	Relatywna cena produktu w odniesieniu do rynku
Smak	Liczba naturalna	1-5	Relatywna jakość smaku produktu w odniesieniu do rynku
Opakowanie	Liczba naturalna	1-5	Relatywna atrakcyjność opakowania produktu w odniesieniu do rynku
Premium	Zmienna binarna	0-1	Czy produkt jest postrzegany jako marka premium (tj. kierownik)
Budżetowy	Zmienna binarna	0-1	Czy produkt jest postrzegany jako marka budżetowa (tj. kierownik)
Lager	Zmienna binarna	0-1	Czy produkt należy do piw typu lager?
Smakowe	Zmienna binarna	0-1	Czy produkt należy do piw smakowych?
Marketing	Liczba naturalna	0-5	Wysokość nakładów na marketing marki

3.5.4 Konkurencja

W założeniach przyjmujemy, że konkurencja jest pasywna - tj. nie podejmuje działań ani decyzji w trakcie trwania symulacji. Wynika to z odmiennego celu badania, którym jest analiza działania algorytmów optymalizacyjnych. Nagłe zmiany sprzedaży spowodowane np. obniżeniem ceny przez konkurencję spowodowałyby wątpliwości interpretacyjne i są zbędne. Konkurencja jest za to potrzebna do stworzenia alternatywnych dla symulowanego produktu, o odmiennych cechach i przyciągających klientów o specyficznych charakterystykach, i jej rola ogranicza się do wprowadzenia go na rynek.

3.5.5 Trasy

Klasa *trasy* przechowuje wszystkie możliwe kombinacje łańcuchów dostaw, składające się z instancji klas *fabryka*, *magazyn*, *sklep* oraz *ścieżek* pomiędzy nimi ⁴². W przypadku

⁴²Jako ścieżkę rozumiemy sekwencję punktów z drogami, jakie trzeba przebyć od fabryki do magazynu i od magazynu do sklepu

złożonych grafów (czyli sytuacji, kiedy przedsiębiorstwo składa się z wielu jednostek) ilość kombinacji może uniemożliwić swobodne przetwarzanie klasy w pamięci komputera, jednak w takim wypadku można predefiniować zbiór możliwych łańcuchów, spośród których model będzie wybierał najbardziej optymalne trasy.

Dla każdej z tras, poza informacjami o jednostkach wchodzących w skład przedsiębiorstwa, przechowujemy także informację o *ścieżkach* pomiędzy jednostkami. Wynika to z faktu, że transporty pomiędzy poszczególnymi jednostkami także mogą doświadczać efektów skali,⁴³ a długość ścieżki bezpośrednio wpływa na koszt łańcucha - koszt wzrasta w stosunku do długości ścieżki.

3.6 Symulowanie decyzji konsumenckich

Jak wskazano na diagramie ??, podczas wizyty każdego z wirtualnych konsumentów w sklepie symulujemy jego decyzję co do zakupu towaru. W tej części chcemy, aby decyzje konsumentów w modelu jak najbardziej przypominały decyzje klientów w analogicznych sytuacjach w świecie rzeczywistym.⁴⁴

Opierając się na Sagan (2011), wiemy, że jesteśmy w stanie zdefiniować kluczowe cechy konsumenta i cechy marketingowe produktu jako zbiór zmiennych ilościowych. Stąd, w grze eksperymentalnej przeprowadzonej na potrzeby pracy⁴⁵, poproszono uczestników o stwierdzenie, jakie produkty z dostępnej listy kupi klient o charakterystyce wylosowanej przez program. Po odpowiedzi udzielonej przez gracza, predefiniowane, jakościowe cechy produktu były transponowane na wartości liczbowe⁴⁶ i wraz z ilościowymi cechami klienta oraz zmienną binarną przechowującą informacje *kupił/niekupił* zapisywany na serwerze SQL.

W programie dane te służą do budowy drzewa klasyfikacyjnego, które - ponieważ dane liczbiowe są w identycznej formie- dla każdej kwerendy o klienta i produkt zwraca prawdopodobieństwo zakupu. Stąd, dla każdego agenta możemy zbudować listę prawdo-

⁴³Przykładowo, pod kątem kosztów na produkt, transport 100 produktów może być bardziej opłacalny niż 10 ze względu na rozłożenie kosztów stałych na większą ilość produktów

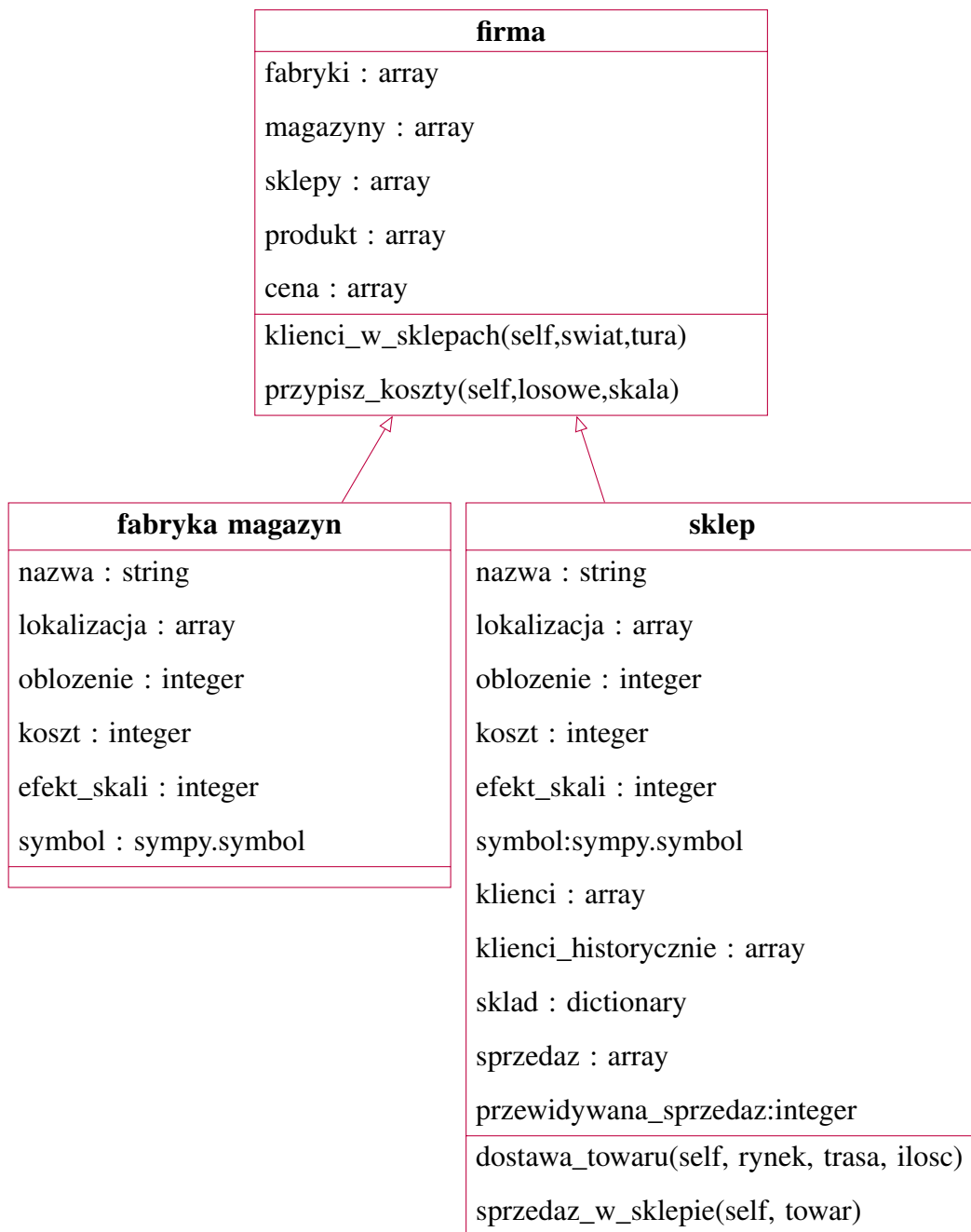
⁴⁴Warto odnotować, że to nie jest to samo co późniejsze przewidywanie "prognozowanej sprzedaży"

⁴⁵Gra dostępna jest pod adresem <http://serwer1418288.home.pl/test/piwo/zapisy.php>

⁴⁶Na przykład, piwo Grolsh jest drogie, klasy premium i jest lager, stąd otrzyma zapis [5,1,1], a tani smakowy Redd's [3,0,0].

podobieństw zakupu każdego z towarów dostępnego na rynku, i w ten sposób losowaniem symulować decyzje konsumenckie. To podejście gwarantuje wysokie podobieństwo z wyborami realnych konsumentów ponieważ

- ze względu na wysoką ilość rekordów jest duża szansa, że istnieje zapis o decyzjach klienta o bardzo podobnej charakterystyce
- baza danych jest generowana przez decyzje ludzkie, zapewniając wysoką zgodność z analogicznymi wyborami w świecie rzeczywistym
- wybór drzewa klasyfikacyjnego jako metody i wysoki n rekordów pozwala na bardzo pożądaną w tej sytuacji *overfit*, który jak opisuje James et al. (2013) powoduje, że dostajemy bardzo dokładne odwzorowanie zbioru uczącego.



Wykres 12: Diagram UML klasy firma, fabryka, magazyn oraz sklep *Źródło:*
opracowanie własne

4 Efekty działania algorytmu optymalizacyjnego

4.1 Założenia modelu

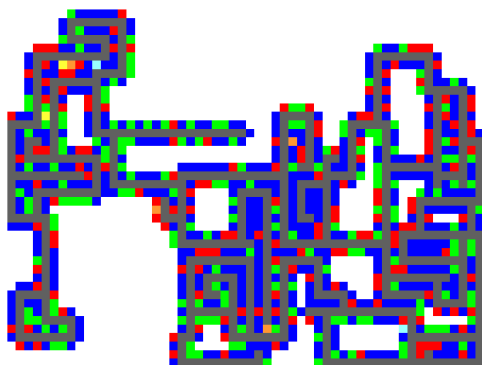
Weryfikacja działania algorytmu odbywać się będzie w modelu opisanym w rozdziale 3, w oparciu o losowo wygenerowane środowisko działania modelu (*świat i rynek*), których właściwości zostały opisane w tabeli 3. Mapy powstałego w ten sposób środowiska zostały przedstawione na wykresach 13a i 13b, a symulowane przedsiębiorstwo zostało przedstawione w postaci grafu na wykresie 16.

Tabela 3: Założenia generowanego świata. *Źródło:* opracowanie własne.

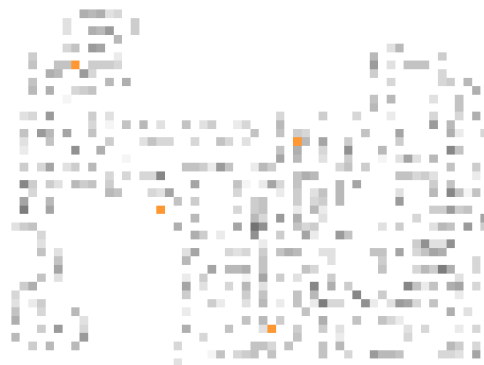
Właściwość	Założenie
Populacja	2500
Wymiar x	60
Wymiar y	60
Udział dróg	0.3
Gęstość zakrętów dróg	0.05
Udział budynków mieszkalnych	0.6
Udział biurowców ⁴⁷	0.2
Udział przestrzeni komercyjnej ⁴⁸	0.2
Ilość fabryk	2
Ilość magazynów	2
Ilość sklepów	4
Cena produktu	7

Konsumenci

Cechy konsumentów zostały wygenerowane zgodnie z rozkładami zawartymi w Narodowym Spisie Ludności (Główny Urząd Statystyczny (2011)) oraz XI Ogólnopolskim Badaniu Wynagrodzeń Sedlak&Sedlak (2013) w celu zapewnienia spójności z rzeczywistą strukturą społeczną. Histogramy cech klientów w modelu zostały zaprezentowane na wykresie 14.



(a) Mapa typów lokacji na mapie



(b) Histogram występowania konsumentów na mapie.

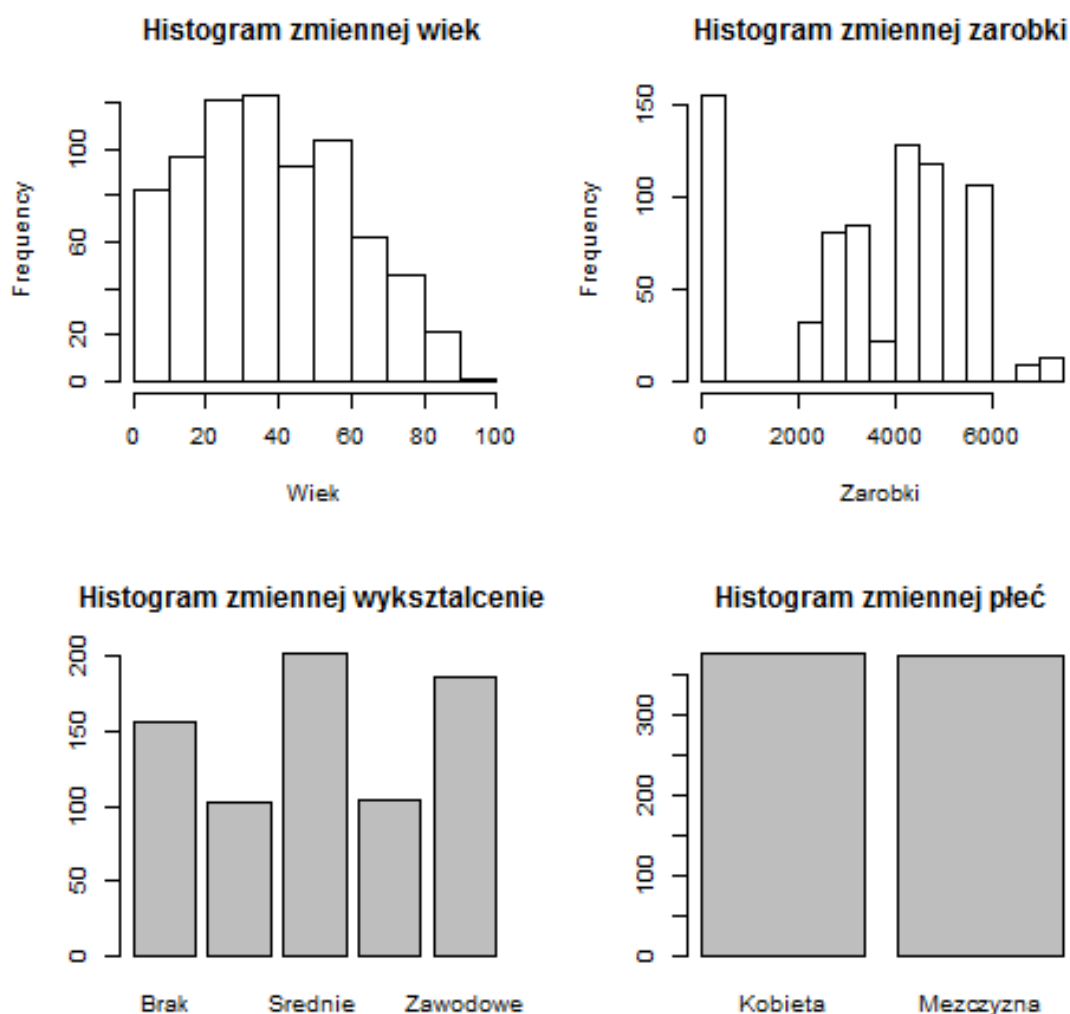
Wykres 13: Mapa wygenerowanych macierzy lokalizacji w modelu

Jak zostało opisane w rozdziale 3.5.1, cechy klientów wpływają na ich decyzji konsumenckie i wybór produktów w sklepie. Dlatego profil klienta konsumującego symulowany produkt (zob. tabela 5) będzie dystynktywny, tak jak w normalnych warunkach rynków różne marki przyciągają klientów o różnych profilach i cechach demograficznych. Właściwość tą wykorzystujemy do modelowania predyktywnego prawdopodobieństwa zakupu przy danej charakterystyce klienta w czasie $t + 1$. Porównanie rozkładów cech wszystkich klientów i konsumentów marki zaprezentowano na wykresie 15.

Przedsiębiorstwo

Symulowane przedsiębiorstwo, zdefiniowane jak w rozdziale 2.2.3, składa się z zaprezentowanej w tabeli 3 ilości jednostek — 2 fabryk, 2 magazynów i 4 sklepów, które wspólnie tworzą $\binom{2}{1} \times \binom{2}{1} \times \binom{4}{1} = 16$ możliwych ścieżek. Lokacje poszczególnych jednostek zostały zaprezentowane na rysunku 17.

Każdy z j elementów przedsiębiorstwa oraz krawędzie d je łączące powiązane jest z funkcją kosztu f_j , które w celu symulacji efektów skali są nieliniowe. Dla fabryk, magazynów i sklepów zakładamy ujemne korzyści skali, podczas gdy dla transportu na krawędziach — dodatnie. Dokładne funkcje kosztów wykorzystane w modelu opisane są w tabeli 4.



Wykres 14: Histogramy cech agentów. *Źródło:* opracowanie własne

Produkty

W modelu zakładamy, że na rynku obecnych jest 7 marek piwa, z których jedno należy do symulowanego przez nas przedsiębiorstwa (produkt o nazwie **Symulowane**). Jak opisano w rozdziale 3.5.3, każde z piw zdefiniowane jest 7 zmiennymi marketingowymi, których cechy produktów zostały zaprezentowane zostały w tabeli 5. Wyniki sprzedaży poszczególnych marek w jednostkach czasu t zaprezentowane są na wykresie 18.

Zauważmy w tabeli 5, że piwo należące do symulowanego przez nas przedsiębiorstwa należy do jednych z najtańszych piw na rynku, jest piwem smakowym o przeciętnych walorach smakowych, ale za to intensywnych nakładach na marketing i dużej uwadze

Tabela 4: Założone w modelu funkcje kosztów. *Źródło:* opracowanie własne.

Jednostka	Symbol	Funkcja kosztu ⁴⁹
Fabryka	$fa \in FA$	$f_{fa} = 1.3 \times x_{fa}^{1.03}$
Magazyn	$ma \in MA$	$f_{ma} = 1.2 \times x_{ma}^{1.05}$
Sklep	$sk \in SK$	$f_{sk} = 1.1 \times x_{sk}^{1.07}$
Droga (krawędź)	$d \in D^{50}$	$f_d = 0.005 \times l_d \times x_d^{0.95}$

poświęconej projektowaniu opakowania. Skonfrontowanie tego z profilem klienta widocznym na wykresie 15 pozwala zrozumieć, dlaczego nabywcy naszego piwa należą do grupy gorzej sytuowanych finansowo, raczej młodszych ⁵¹, z dominującą grupą kobiet. Profil wykształcenia nie różni się znacząco od przeciętnej, z delikatnie mniejszym udziałem klientów o wyższym wykształceniu.

Tabela 5: Cechy symulowanych produktów. ⁵² ⁵³*Źródło:* opracowanie własne.

	Cena	Smak	Opakowanie	P	B	L	S	Marketing
Symulowane	1	3	5	0	0	0	1	4
Slaskie	4	4	2	1	0	1	0	4
Lebskie	4	2	2	1	0	0	1	5
Babskie	4	3	5	0	1	0	1	5
Pszczeniczne	2	2	2	0	0	0	1	2
Opolskie	4	3	5	0	0	0	1	4
Mocne	4	2	2	1	0	1	0	3

⁵¹ Wynika to z połączenia faktów, że młode osoby mniej zarabiają, oraz piwa smakowe nie są zbyt popularne wśród starszych roczników

4.2 Przewidywanie decyzji konsumentów

Przewidywanie decyzji konsumentów odbywa się zgodnie z założeniami przedstawionymi w rozdziale 2.2.2.⁵⁴. Kolejne kroki prognozowania sprzedaży w każdym ze sklepów w czasie $t + 1$ to

1. Na podstawie historii transakcji obliczamy średnią ilość klientów n odwiedzających sklep w jednostce czasu t
2. Na podstawie historii cech klientów kupujących produkt, obliczamy prawdopodobieństwo warunkowe wystąpienia wszystkich zestawów cech, i losujemy cechy dla n klientów w $t + 1$
3. Dla każdego z tak otrzymanych klientów na podstawie przeszłej historii wyborów klientów, wykorzystując metodę regresji logistycznej bądź klasyfikacji k-means, obliczamy prawdopodobieństwo zakupu produktu symulowanej firmy
4. Kroki 1-3 powtarzamy założoną ilość razy⁵⁵ i obliczamy średnią z trzech obserwacji

Analizując wyniki prognozowania, przedstawione na wykresach na wykresach ?? możemy zobserwować różnicę w skuteczności prognoz obu stosowanych metod — *regresji logistycznej* oraz *k-nearest neighbours*. Chociaż współczynnik determinacji R^2 jest wyższy dla metody regresji logistycznej ($R_{lg}^2 = 0,7912$) niż dla K-nearest neighbours ($R_{kn}^2 = 0,7591$), to najcenniejszą cechą regresji jest, że nie przeszacowuje sprzedaży tak jak to robi K-nearest neighbours. Nadmierny optymizm metody K-nearest neighbours powoduje, że zamawia się produkty o których wiemy, że nie zostaną sprzedane, ponosząc niepotrzebne koszty produkcji i transportu (zob. wykres ??).

⁵⁴W modelu stosowane jest pewne uproszczenie podczas prognozowania ilości klientów w sklepie w $t + 1$. Ponieważ wizyta klienta w sklepie jest w modelu zdarzeniem losowym (losuje się jedna z czterech lokalizacji docelowych), a model nie symuluje efektów zewnętrznych jak pogoda czy dzień tygodnia, do przewidywania liczby klientów odwiedzających sklep w chwili $t + 1$ stosowana jest średnia. Brak jest bowiem zmiennych objaśniających które mogłyby zostać wykorzystane w modelu, jednak ich wprowadzenie dodawałoby niepotrzebną złożoność do modelu, a zastosowane uproszczenie nie ma wpływu na badanie wyników działania algorytmu

⁵⁵W modelu przyjmujemy 3 iteracje

Wynika to z faktu, że K-nearest neighbours dla współczynnika $k = 3$ ma większe trudności z jednoznacznym określeniem klasy przewidywanej zmiennej ⁵⁶ niż regresja logistyczna, jak możemy to zaobserwować na wykresie 20. Regresja logistyczna jest o wiele bardziej skłonna to przyporządkowania niskiego prawdopodobieństwa zakupu, jednak z większą pewnością określa klientów którzy kupią produkt. Tymczasem, k-nearest neighbours często wskazuje wartość prawdopodobieństwa w środku przedziału $(0, 1)$, dając odpowiedź niejednoznaczną. Możliwe, że manipulacja współczynnikiem k wpłynęła by na zachowanie metody, jednak w przypadku modelu za bardziej skuteczną będziemy uważać metodę regresji logistycznej.

4.3 Wpływ działania algorytmu na działanie przedsiębiorstwo

Na początku symulacji przez 20 jednostek czasu t , w celu uzyskania punktu odniesienia, przedsiębiorstwo podejmuje decyzje na podstawie predefiniowanych zasad opisanych w rozdziale 3.1, czyli

- towar zamawiamy jest zawsze z najbliższego magazynu
- ilość zamówionego towaru do każdego ze sklepów równa jest sprzedaży t_0 ⁵⁷

Po 20 jednostkach czasu na podstawie dotychczasowej historii transakcji inicjalizowane są modele regresji logistycznej oraz k-nearest neighbours, które przez następne 10 jednostek czasu t służą do prognozowania sprzedaży w każdym ze sklepów w $t + 1$. Zgodnie z przyjętymi w rozdziale 2.2.3, wiedza o sprzedaży w każdym ze sklepów w $t + 1$ pozwala nam, po odpowiednich przekształceniach, zbudować algorytm optymalizujący procesy produkcyjno-logistyczne, przede wszystkim pod kątem alokacji wolumenów produkcji i tras dostaw pomiędzy poszczególne jednostki przedsiębiorstwa.

⁵⁶Warto przypomnieć, że w tym momencie modelujemy to, czy dany klient będzie w klasie 0 (*nie kupił*) czy 1 (*kupił*)

⁵⁷Gdzie t_0 to runda próbna, która ma na celu sprawdzenie popytu na rynku i nie jest zapisywana do wyników firmy. Jej wdrożenie ma na celu odzwierciedlenie w modelu wiedzy powszechnej o rynku, tj. przedsiębiorcy wiedzą ile mogą sprzedać na podstawie tego, co sprzedawali w niedalekiej przeszłości albo na podstawie raportów rynkowych

Alokacje na trasach dostaw (ścieżkach)

W rezultacie działania modelu otrzymujemy wektor A , który dla każdego elementu zbioru $w \in W$ (zob. rozdział 2.2.3) zawiera optymalny ⁵⁸ wolumen dostaw w $t + 1$. Wektory otrzymane w toku działania modelu zostały zaprezentowane na wykresie 21.

Analiza wykresu wyraźnie pokazuje, że przedsiębiorstwo sterowane przez algorytm jest o wiele bardziej elastyczne co do wyboru ścieżek, na co wskazuje rosnąca po $t = 20$ wariancja alokacji. Ponadto, algorytm potrafi dostarczyć towar do sklepu z wielu źródeł. Biorąc pod uwagę, że funkcje kosztów są nieliniowe a we wszystkich jednostkach firmy występują negatywne efekty skali, pozwala to znaczące ograniczenie kosztów.

To o tyle ważna obserwacja, że w praktyce biznesowej podobne metody ograniczenia kosztów nie są szeroko stosowane. Wręcz przeciwnie, często pracownicy firmy podczas składania zamówień podejmują decyzje, które lokalnie wydają się im być najlepsze (np. zamawianie z najbliższego magazynu), ale z punktu widzenia całego systemu nie są optymalne. Zastosowanie podobnych do opisywanego algorytmów może przynieść znaczące oszczędności w przedsiębiorstwach.

Przychody i koszty

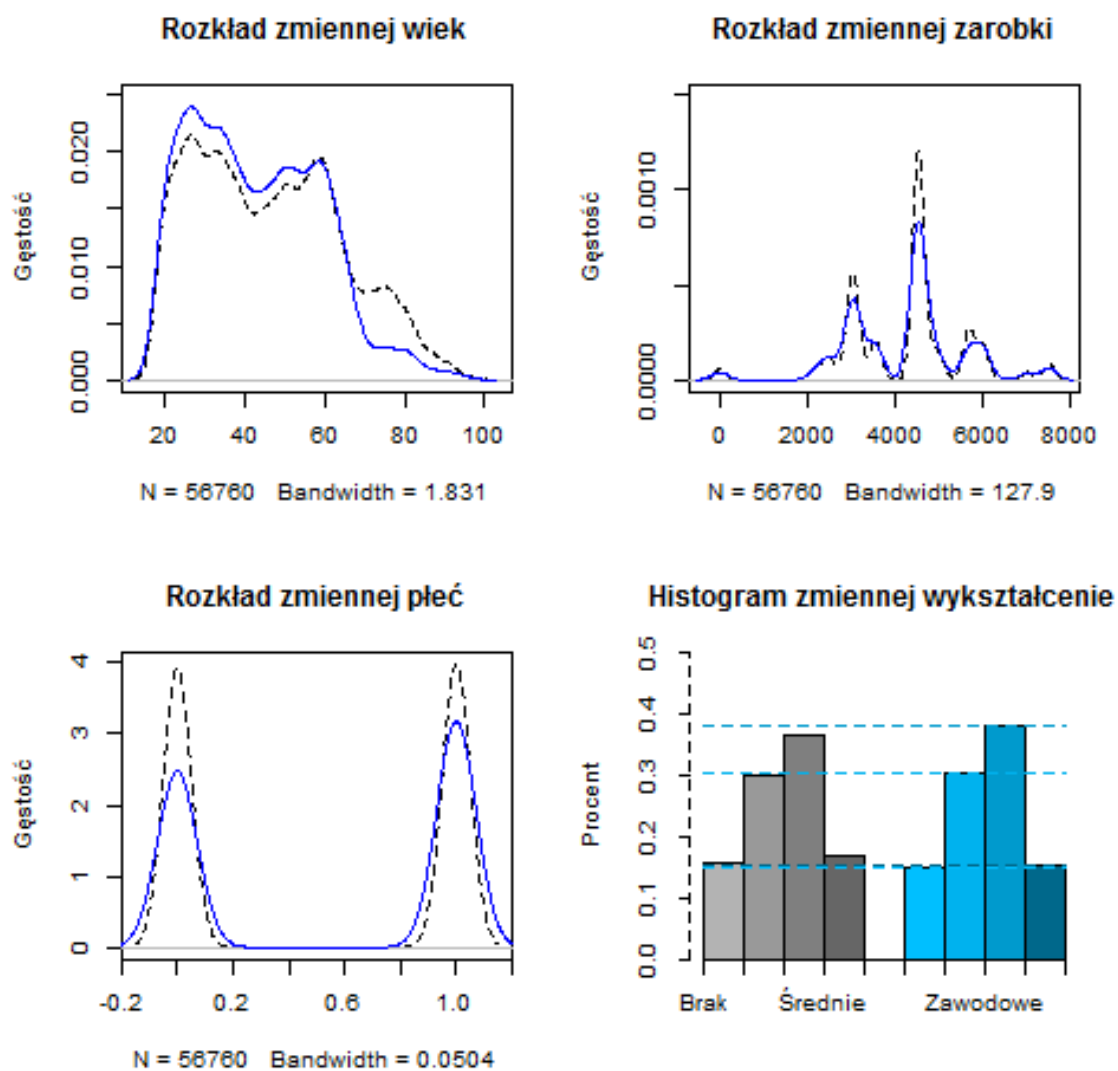
Obserwacja przychodów, kosztów i zysków przedsiębiorstwa w modelu, przedstawione odpowiednio na wykresach 23b oraz 23a, pozwala zauważyć, że działanie algorytmu wpłynęło *in plus* na wyniki firmy. Warto odnotować, że przedsiębiorstwo w początkowych turach miało stabilne koszty, które jak wspomniano w rozdziale 3.1, starano się minimalizować, zamawiając towary tylko z najbliższego magazynu.

Chociaż jest to racjonalna strategia, nie była jednak skuteczna w obliczu dużej zmienności ilości klientów w sklepach, którą można zaobserwować na wykresie 19, była wysoce nieskuteczna — zyskowe okresy przeplatały się z nierentownym. Straty w niektórych okresach wynikały przede wszystkim z niepotrzebnie wysokich kosztów zamówień oraz nieumiejętności dostosowania się do wysokiej zmienności wolumenu sprzedaży, co wymagało większej elastyczności w zamówieniach. Model, jak zaobserwowano w 4.3

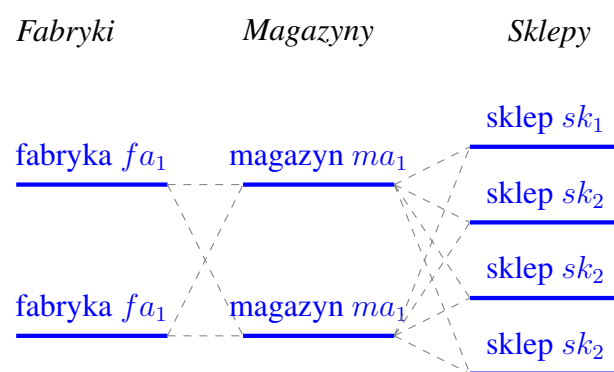
⁵⁸Zastrzeżenia wobec optymalności alokacji sugerowanych przez algorytm znalazły się w rozdziale 2.3

jest bardziej elastyczny, dzięki czemu mógł lepiej dostosowywać się do popytu i odnotowywać zysk w każdej turze.

Dodatkowo, gdyby zmienność nie była wywołana losowo, jak w modelu, ale zależna od serii czynników (pogoda, dzień tygodnia, etc), możliwe byłoby stworzenie modelu który wyłapywałby takie zmiany w tendencjach i potrafiłby dostosować do nich produkcję. Dobrym przykładem skuteczności takiego algorytmu byłby model przewidujący sprzedaż lodów, który przy prognozie dobrej pogody zwiększałby zapasy, a przy ochłodzeniach reduktował produkcję. Zazwyczaj takie procesy koordynowane są przez ludzi, jednak duża ilość zmiennych oraz możliwa ich heterogeniczność (różna pogoda w różnych częściach Polski) powoduje, że pewne prognozy są zbyt złożone żebyśmy mogli je przewidywać z wysoką dokładnością bez metod matematycznych.



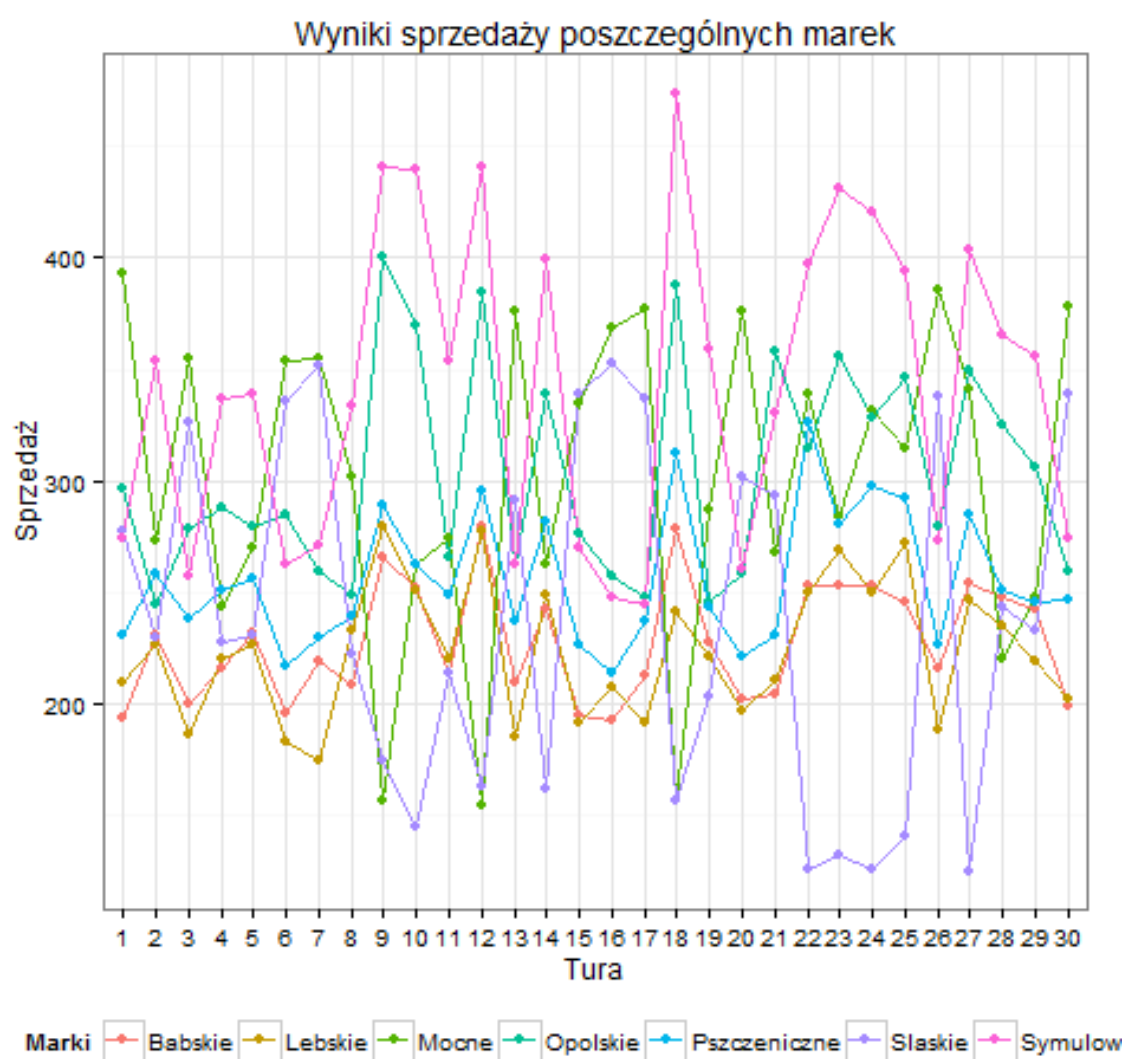
Wykres 15: Cechy konsumentów marki w porównaniu do wszystkich agentów. Na czarno zaznaczone dane dla wszystkich klientów, na niebiesko — dla klientów marki. *Źródło:* opracowanie własne



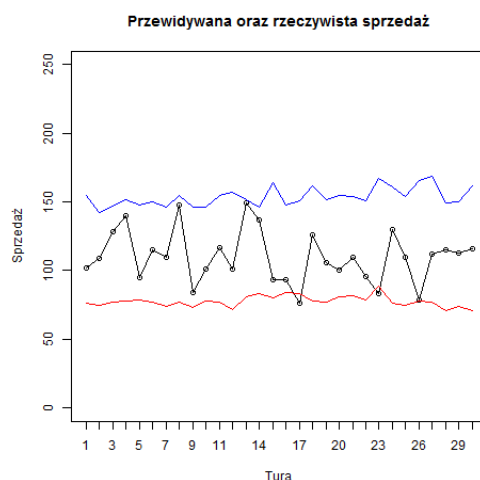
Wykres 16: Modelowane przedsiębiorstwo przedstawione jako graf kierunkowy Źródło: opracowanie własne



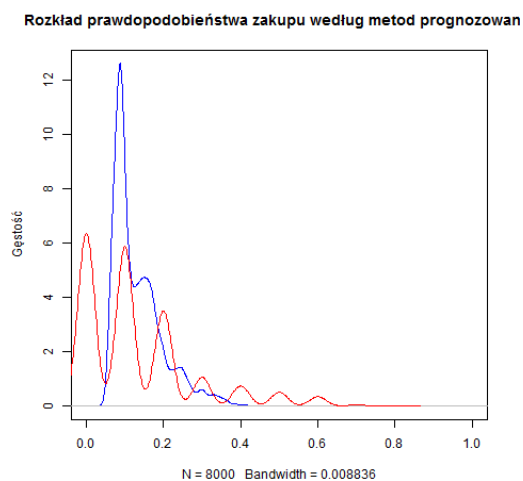
Wykres 17: Mapa z zaznaczonymi jednostkami przedsiębiorstwa. Źródło:
opracowanie własne



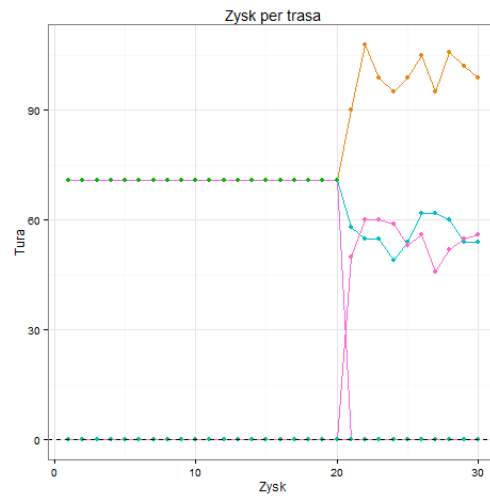
Wykres 18: Statystyki sprzedaży marek symulowanych w modelu. *Źródło:* opracowanie własne



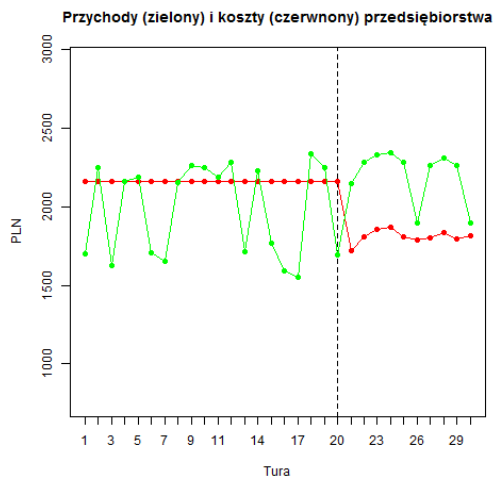
Wykres 19: Przewidywana i realna sprzedaż w modelu. Prognoza metodą regresji logistycznej zaznaczona na niebiesko, metoda K-nearest neighbours — czerwono *Źródło:* opracowanie własne



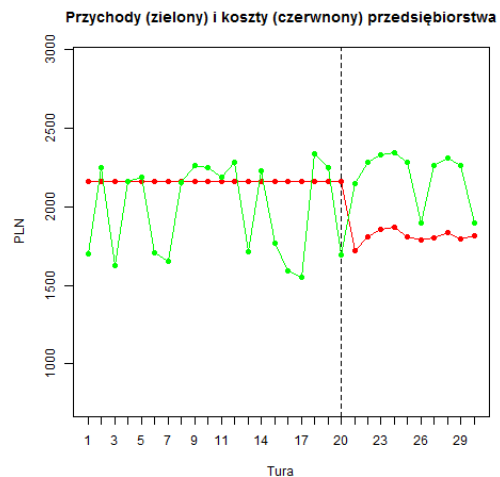
Wykres 20: Rozkład prawdopodobieństwa zakupu w prognozach według metod prognoz. Prognoza metodą regresji logistycznej zaznaczona na niebiesko, metoda K-nearest neighbours — czerwono *Źródło:* opracowanie własne



Wykres 21: Alokacje wolumenów dostaw wśród możliwych ścieżek. Przerywana liniia oznacza inicjalizację działania modelu. Źródło: opracowanie własne

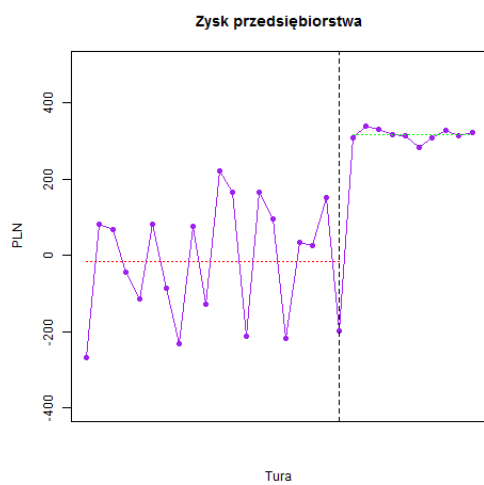


(a) Przychody i koszty symulowane przedsiębiorstwa w symulacji z metodą prognozowania regresji logistycznej

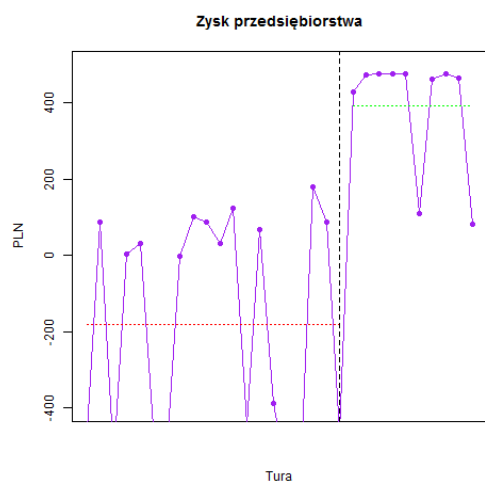


(b) Przychody i koszty symulowane przedsiębiorstwa w symulacji z metodą prognozowania k-nearest neighbours

Wykres 22: Wyniki finansowe symulowanego przedsiębiorstwa. Przerywana linia oznacza inicjalizację algorytmu. Źródło: opracowanie własne



(a) Zysk symulowanego przedsiębiorstwa, z zaznaczonymi średnimi zyskami przed i po inicjalizacji algorytmu



(b) Zysk symulowanego przedsiębiorstwa w symulacji z metodą prognozowania regresji logistycznej, z zaznaczonymi średnimi zyskami przed i po inicjalizacji algorytmu

Wykres 23: Zysk symulowanego przedsiębiorstwa, z zaznaczonymi średnimi zyskami przed i po inicjalizacji algorytmu. Przerywana linia oznacza inicjalizację algorytmu. *Źródło:* opracowanie własne

Literatura

- Brynjolfsson, E., Hitt, L. M. i Kim, H. H. (2011), 'Strength in Numbers: How does data-driven decision-making affect firm performance?', *ICIS 2011 Proceedings* p. 18.
- Buckinx, W., Verstraeten, G. i den Poel, D. V. (2007), 'Predicting customer loyalty using the internal transactional database', *Expert Systems with Applications* **32**(1), 125 – 134.
URL: <http://www.sciencedirect.com/science/article/pii/S0957417405003143>
- Davenport, T., D'Amboise, L. i Lucker, J. (2011), 'Know what your customers want before they do', *Harvard Business Review* .
- Frey, C. B. i Osborne, M. (2013), 'The future of employment: How susceptible are jobs to computerisation?', Online PDF.
URL: <http://www.oxfordmartin.ox.ac.uk/publications/view/1314>
- Główny Urząd Statystyczny (2011), Narodowy spis ludności 2011, Technical report, Główny Urząd Statystyczny.
- Hastie, T., Tibshirani, R. i Friedman, J. (2001), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA.
- James, G., Witten, D., Hastie, T. i Tibshirani, R. (2013), *An Introduction to Statistical Learning with Applications in R*, Springer-Verlag New York.
- Kaminski, B. (2012), *Podejście wieloagentowe do modelowania rynków*, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.
- Kawa, A. (2010), 'System wieloagentowy do budowania łańcuchów dostaw w branży komputerowej', *Zarządzanie Przedsiębiorstwem* **13**(1), 26–34.
- McKinsey Global Institute (2011), Big data: The next frontier for innovation, competition, and productivity, Technical report, McKinsey Global Institute.
- Moyaux, T., Chaib-Draa, B. i D'Amours, S. (2006), 'Supply chain management and multiagent systems: An overview', *Multiagent based Supply Chain Management* **4**, 1–27.

- Porter, M. E. (1985), *Competitive advantage: Creating and sustaining superior performance*, Free Press, New York and London.
- PwC (2014), CEO Survey 2014, Technical report.
- Sagan, A. (2011), 'Modele strukturalne w analizie zachowan konsumenta - ewolucja podejsc', *Konsumpcja i rozwoj*.
- Schelling, T. (1971), 'Dynamic models of segregation', *Journal of Mathematical Sociology*.
- Sedlak&Sedlak (2013), Xi ogolnopolskie badanie wynagrodzen 2013, Technical report, Sedlak&Sedlak.
- The Economist (2010), Data, data everywhere, Technical report.
- Wooldridge, M. i Jennings, N. R. (1995), 'Intelligent agents: theory and practice', *The Knowledge Engineering Review* **10**, 115–152.

Spis tablic

1	Zastosowane biblioteki języka Python	21
2	Cechy charakteryzujące produkt w modelu	31
3	Właściwości generowanego świata	35
4	Założone w modelu funkcje kosztów	38
5	Cechy symulowanych produktów	38

Spis rysunków

1	Przedsiębiorstwo jako graf kierunkowy	8
2	Rozważane przedsiębiorstwo przedstawione jako graf kierunkowy	13
3	Proponowany algorytm optymalizacyjny	16
4	Struktura działania programu	22
5	Diagram UML klasy rynek	23
6	Diagram UML klasy świat, lokalizacja i konsument	25
7	Algorytm generowania świata w modelu	26
8	Przykładowa wygenerowana mapa	27
9	Algorytm wyszukiwania drogi	28
10	Diagram UML klasy konsument	29
11	Rozkłady zmiennych cech klasy konsument	30
12	Diagram UML klasy firma, fabryka, magazyn oraz sklep	34
13	Krótką nazwa II	36
14	Histogramy cech agentów	37
15	Cechy konsumentów marki w porównaniu do wszystkich agentów	43
16	Modelowane przedsiębiorstwo przedstawione jako graf kierunkowy	44
17	Krótką nazwa X	45
18	Statystyki sprzedaży marek symulowanych w modelu	46
19	Przewidywana i realna sprzedaż w modelu	47
20	Rozkład prawdopodobieństwa zakupu w prognozach	47
21	Alokacje wolumenów dostaw wśród możliwych ścieżek	48
22	Wyniki finansowe symulowanego przedsiębiorstwa	48
23	Zysk symulowanego przedsiębiorstwa	49

Streszczenie