

# Drzewa decyzyjne i wnioskowanie

Mateusz Jakubczak, Karol Oleszek, Krzysztof Olipra

## Zbiór danych

Badany zbiór danych opisuje przesłanki i decyzję czy pracować w danej firmie informatycznej.

## Przesłanki

czy\_zagraniczna 0/1 (polska firma, zagraniczna firma)

mozliwosci\_rozwoju 0/1/2 (małe, średnie, duże)

wielkosc\_firmy 0/1/2 (mała firma, średnia firma, duża firma)

typ\_firmy 0/1/2 (startup, korporacja, firma państwowa)

wynagrodzenie 0/1/2/3 (brak, małe, średnie, duże)

## Decyzja

czy\_pracowac 0/1/2 (nie, może, tak)

## Dane

Decyzje zostały wygenerowane dla całego iloczynu kartezyjskiego przesłanek na podstawie poniższych reguł użytych w podanej kolejności:

1. Domyślnie może pracuj
2. Jeżeli firma zagraniczna to pracuj
3. Jeżeli mała firma to nie pracuj
4. Jeżeli brak wynagrodzenia, ale duże możliwości rozwoju to pracuj
5. Jeżeli polski startup to nie pracuj
6. Jeżeli firma państwowa z wysokim wynagrodzeniem to pracuj

## Algorytm

Użyty algorytm do stworzenia to iteratywny dychotomizator 3 (ID3) autorstwa Quinlana. Opiera się on na maksymalizacji zysku informacyjnego.

## Uzyskane drzewo decyzyjne:

Pytanie: Czy wielkosc\_firmy równe jest mała?

Tak:

| Pytanie: Czy typ\_firmy równe jest startup?

| Tak:

|| Pytanie: Czy wynagrodzenie równe jest brak?

|| Tak:

||| Pytanie: Czy czy\_zagraniczna równe jest nie?

||| Tak:

|||| Decyzja: Nie pracuj.

|||| Nie:

||||| Pytanie: Czy mozliwosci\_rozwoju równe jest małe?

||||| Tak:

||||| Decyzja: Nie pracuj.

||||| Nie:

||||| Pytanie: Czy mozliwosci\_rozwoju równe jest średnie?

||||| Tak:

||||| Decyzja: Nie pracuj.

||||| Nie:

||||| Decyzja: Pracuj.

|| Nie:

|| Decyzja: Nie pracuj.

| Nie:

| Pytanie: Czy typ\_firmy równe jest korporacja?

| Tak:

|| Pytanie: Czy wynagrodzenie równe jest brak?

|| Tak:

||| Pytanie: Czy mozliwosci\_rozwoju równe jest małe?

||| Tak:

||| Decyzja: Nie pracuj.

||| Nie:

||| Pytanie: Czy mozliwosci\_rozwoju równe jest średnie?

||| Tak:

||| Decyzja: Nie pracuj.

||| Nie:

||| Decyzja: Pracuj.

|| Nie:

|| Decyzja: Nie pracuj.

| Nie:

| Pytanie: Czy wynagrodzenie równe jest średnie?

| Tak:

| Decyzja: Pracuj.

| Nie:

|| Pytanie: Czy wynagrodzenie równe jest małe?

|| Tak:

|| Decyzja: Nie pracuj.

|| Nie:

||| Pytanie: Czy wynagrodzenie równe jest brak?

||| Tak:

||| Pytanie: Czy mozliwosci\_rozwoju równe jest małe?

||| Tak:

||| Decyzja: Nie pracuj.

||| Nie:

||| Pytanie: Czy mozliwosci\_rozwoju równe jest średnie?

||| Tak:

||| Decyzja: Nie pracuj.

||| Nie:

||| Decyzja: Pracuj.

|||| Nie:  
|||| Decyzja: Pracuj.  
Nie:  
| Pytanie: Czy typ\_firmy równe jest startup?  
| Tak:  
| Pytanie: Czy czy\_zagraniczna równe jest nie?  
| Tak:  
|| Decyzja: Nie pracuj.  
|| Nie:  
|| Decyzja: Pracuj.  
| Nie:  
| Pytanie: Czy czy\_zagraniczna równe jest nie?  
| Tak:  
| Pytanie: Czy typ\_firmy równe jest korporacja?  
| Tak:  
|| Pytanie: Czy wynagrodzenie równe jest brak?  
|| Tak:  
|||| Pytanie: Czy mozliwosci\_rozwoju równe jest małe?  
|||| Tak:  
|||| Decyzja: Może pracuj  
|||| Nie:  
|||| Pytanie: Czy mozliwosci\_rozwoju równe jest średnie?  
|||| Tak:  
|||| Decyzja: Może pracuj  
|||| Nie:  
|||| Decyzja: Pracuj.  
||| Nie:  
||| Decyzja: Może pracuj  
|| Nie:  
|| Pytanie: Czy wynagrodzenie równe jest średnie?  
|| Tak:  
|| Decyzja: Pracuj.  
|| Nie:  
|| Pytanie: Czy wynagrodzenie równe jest małe?  
|| Tak:  
|| Decyzja: Może pracuj  
|| Nie:  
|||| Pytanie: Czy wynagrodzenie równe jest brak?  
|||| Tak:  
|||| Pytanie: Czy mozliwosci\_rozwoju równe jest małe?  
|||| Tak:  
|||| Decyzja: Może pracuj  
|||| Nie:  
|||| Pytanie: Czy mozliwosci\_rozwoju równe jest średnie?  
|||| Tak:  
|||| Decyzja: Może pracuj  
|||| Nie:  
|||| Decyzja: Pracuj.  
|||| Nie:  
|||| Decyzja: Pracuj.  
|| Nie:  
|| Decyzja: Pracuj.

## Podsumowanie

Uzyskane drzewo decyzyjne poprawnie przewiduje konkluzje dla każdego przykładu. Uzyskana struktura jest bardzo rozbudowana; dla bardziej praktycznego problemu należałoby rozważyć przycięcie drzewa w celu lepszej generalizacji.

# Wnioskowanie

## Rodzaje wnioskowania

- W przód (forward chaining)
- Wstecz (backward chaining)
- Mieszane (mixed chaining)

## Wnioskowanie w przód

Wnioskowanie w przód polega na iteracyjnym zbadaniu każdej reguły. W przypadku stwierdzenia poprawności przesłanek, bazę faktów należy powiększyć o fakt z nich wynikający. Proces ten należy powtarzać do momentu aż wszystkie reguły zostaną potwierdzone, uzyskana zostanie wartość celu lub do braku możliwości potwierdzenia pozostałych reguł.

## Wnioskowanie wstecz

Wnioskowanie wstecz polega na wykazaniu prawdziwości hipotezy głównej na podstawie prawdziwości przesłanek użytych w części warunkowej. W przypadku nieznaney wartości logicznej przesłanki należy potraktować ją jako nową hipotezę. Polega ono zatem na sprawdzeniu rekurencyjnie poprawność przesłanek (i ich przesłanek). Proces ten zostaje zakończony w momencie znalezienia hipotezy, dla której wykazano prawdziwość wszystkich przesłanek.

## Wnioskowanie mieszane

Wnioskowanie mieszane jest połączeniem poprzednich dwóch wnioskowań. Najpierw przeprowadzane jest wnioskowanie wstecz, w przypadku natrafienia na przesłankę, której nie da się udowodnić, zostaje przeprowadzone wnioskowanie w przód dla reszty reguł.

## Możliwości rozwoju

mozliwosci\_rozwoju 0/1/2 (małe, średnie, duże)

Na podstawie

- Stare technologie (tak, nie)
- Rodzaj projektu (legacy, nowy, do napisania)
- Możliwość zmiany teamu (nie, tylko w wyjątkowych sytuacjach, często)

Przy regułach

1. Jeżeli **stare technologie = tak** oraz **rodzaj projektu = legacy** lub **rodzaj projektu = nowy** to **małe możliwości rozwoju**
2. Jeżeli **stare technologie = tak** oraz **rodzaj projektu = do napisania** to **średnie możliwości rozwoju**
3. Jeżeli **stare technologie = nie** oraz **rodzaj projektu = nowy** lub **rodzaj projektu = do napisania** to **szanse dużego rozwoju**
4. Jeżeli **szanse dużego rozwoju** oraz **możliwość zmiany teamu = tylko w wyjątkowych sytuacjach lub często** to **duże możliwości rozwoju**
5. Jeżeli **szanse dużego rozwoju** oraz **możliwość zmiany teamu = nie** to **średnie możliwości rozwoju**

# Wynagrodzenie

wynagrodzenie 0/1/2/3 (brak, małe, średnie, duże)

Na podstawie

- Duże miasto (tak/nie) – miasto powyżej 500 tys mieszkańców
- Rodzaj stanowiska (tester, frontend, backend)
- Doświadczenie (junior, mid, senior)

Przy regułach

1. Jeżeli **doświadczenie** = **junior** oraz **rodzaj stanowiska** = **frontend lub tester** to **brak wynagrodzenia**
2. Jeżeli **doświadczenie** = **junior** oraz **rodzaj stanowiska** = **backend** to **stażysta na backendzie**
3. Jeżeli **duże miasto** = **tak** oraz **stażysta na backendzie** to **małe wynagrodzenie**
4. Jeżeli **duże miasto** = **nie** oraz **stażysta na backendzie** to **brak wynagrodzenia**
5. Jeżeli **doświadczenie** = **mid** oraz **rodzaj stanowiska** = **tester** to **mid tester**
6. Jeżeli **duże miasto** = **nie** oraz **mid tester** to **małe wynagrodzenie**
7. Jeżeli **duże miasto** = **tak** oraz **mid tester** to **średnie wynagrodzenie**
8. Jeżeli **doświadczenie** = **mid** oraz **rodzaj stanowiska** = **frontend lub backend** to **średnie wynagrodzenie**
9. Jeżeli **doświadczenie** = **senior** oraz **duże miasto** = **tak** to **duże wynagrodzenie**
10. Jeżeli **doświadczenie** = **senior** oraz **duże miasto** = **nie** to **średnie wynagrodzenie**