

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 kartezyjańskiego 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.