

# **Wczytanie i analiza danych**

# Wstęp

- dane z konkursu KDD Mining Cup 1999
- Zbiór danych zawiera:
  - 8 050 290 rekordów (1 173 MB)
  - 4 940 000 rekordów treningowych (734 MB)
  - 3 110 290 rekordów testowych (430 MB)

# Po eksploracji danych

- zbiór uczący 29087 rekordów
- zbiór testowy 7655 rekordów

# Przed pracą z danymi

- Java 64bit
- Java Environmental Variables -> Runtime Parameters: -Xmx3072m

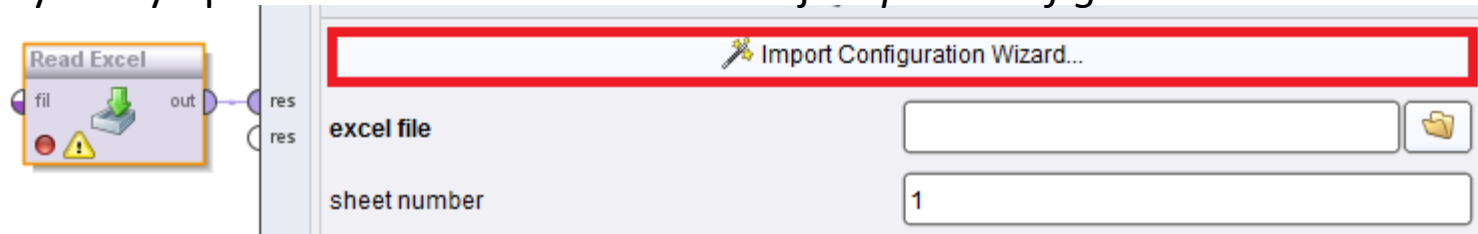
User System					
Platform	Product	Location	Path	Runtime Parameters	Enabled
1.7	1.7.0_45	http://java.s...	D:\Java\jre7\bin\java...	-Xmx3072m	<input checked="" type="checkbox"/>

# Przed pracą z danymi

- Ustawienie maksymalnej pamięci w RapidMinerGUI.bat
- `if "%MAX_JAVA_MEMORY%"==" " set MAX_JAVA_MEMORY=3072`

# Wczytanie danych do RapidMinera

1. Używamy operatora *Read Excel* oraz funkcji *Import Configuration Wizard*



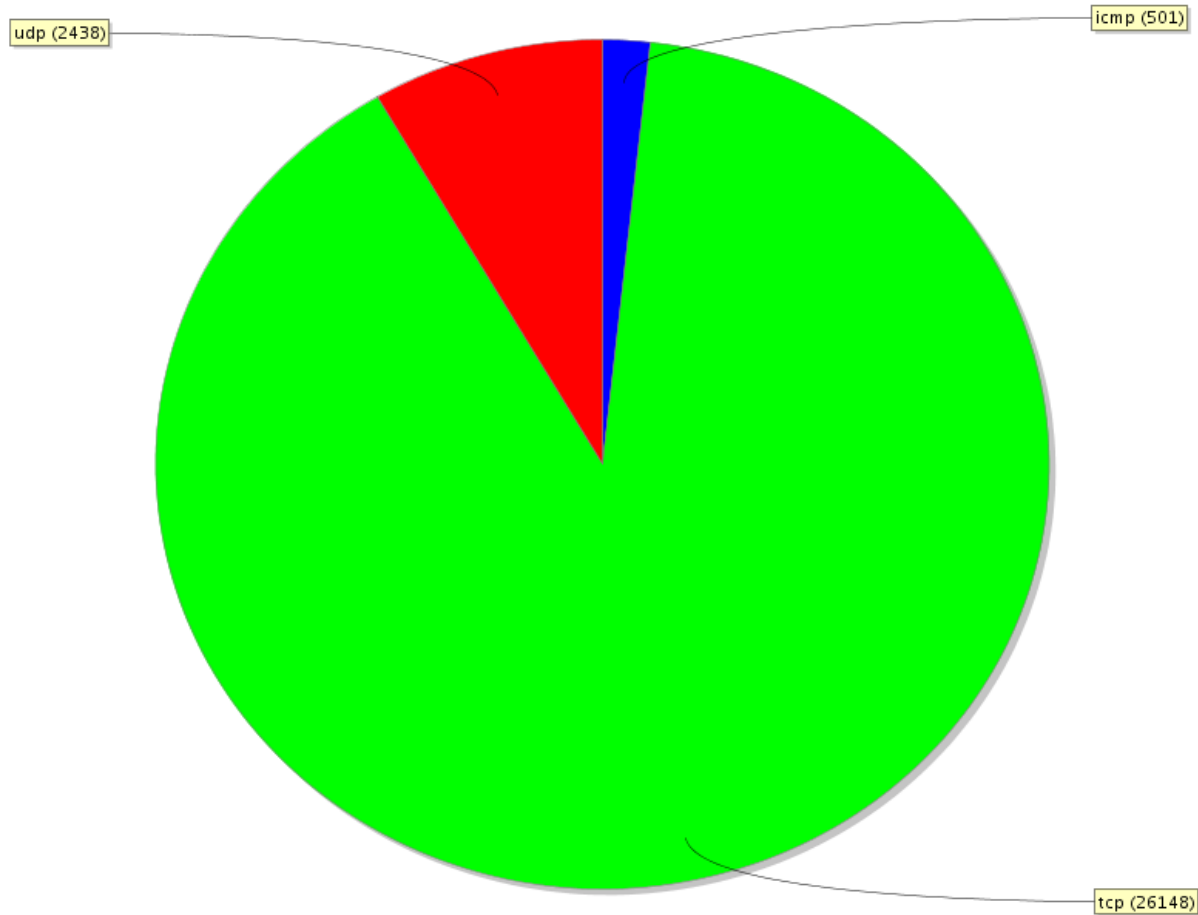
2. Wybieramy plik i przechodzimy do kroku 4 gdzie zmieniamy wartość typu *binominal* typ na *text*

3. *Finish*

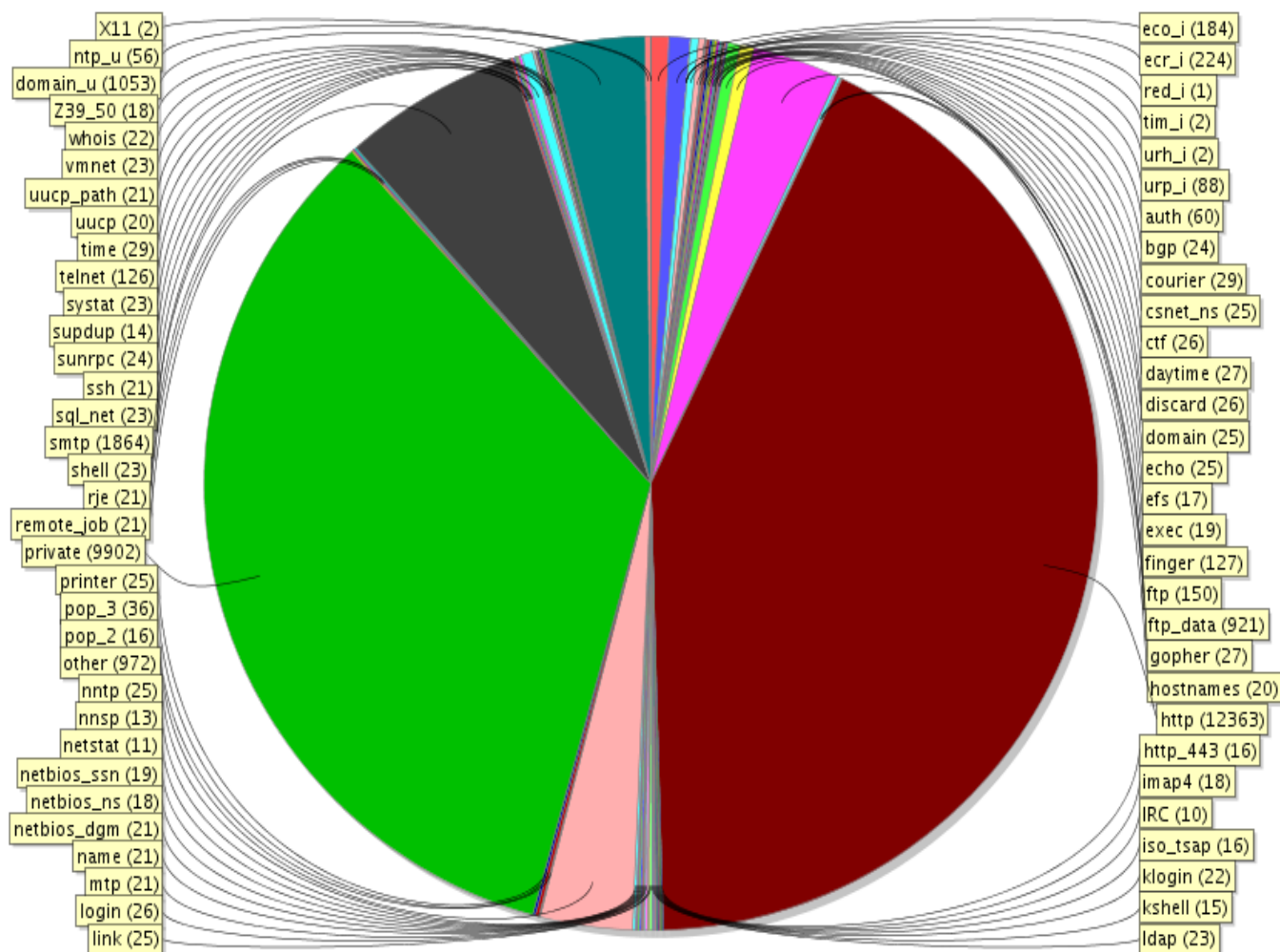
4. Wczytanie danych do pamięci -> 32s.

# Protokoły w zbiorze uczącym

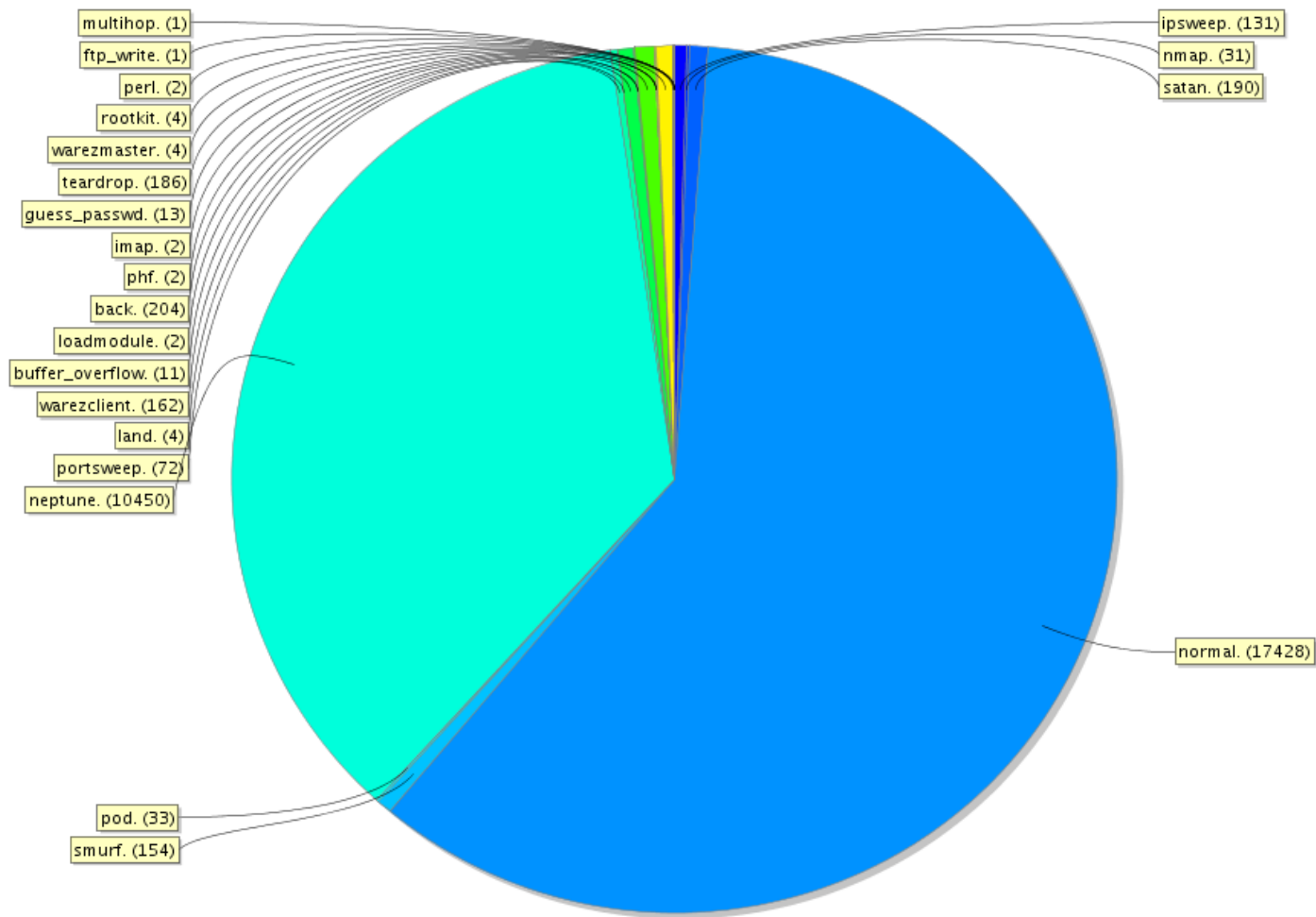
● icmp (501) ● tcp (26148) ● udp (2438)



eco\_i (184) ecr\_i (224) red\_i (1) tim\_i (2) urh\_i (2) urp\_i (88) auth (60) bgp (24) courier (29) csnet\_ns (25) ctf (26) daytime (27) discard (26) domain (25) echo (25) efs (17) exec (19) finger (127) ftp (150) ftp\_data (921) gopher (27) hostnames (20) http (12363) http\_443 (16) imap4 (18) IRC (10) iso\_tsap (16) klogin (22) kshell (15) ldap (23) link (25) login (26) mtp (21) name (21) netbios\_dgm (21) netbios\_ns (18) netbios\_ssn (19) netstat (11) nnsdp (13) nntp (25) other (972) pop\_2 (16) pop\_3 (36) printer (25) private (9902) remote\_job (21) rje (21) shell (23) smtp (1864) sql\_net (23) ssh (21) sunrpc (24) supdup (14) systat (23) telnet (126) time (29) uucp (20) uucp\_path (21) vmnet (23) whois (22) Z39\_50 (18) domain\_u (1053) ntp\_u (56) X11 (2)







# Zależność między rodzajem ataku a protokołem

Atak\Protokół	TCP	UDP	ICMP
DOS	51.42	35	61.90
U2R	8.53	5	0
R2L	11.61	0	0
PROBE	28.44	60	38.09

