УДК [629.735.33(043.2)](http://library.kpi.kharkov.ua/scripts/irbis64r_01/cgiirbis_64.exe?Z21ID=&I21DBN=BOOK&P21DBN=BOOK&S21STN=1&S21REF=5&S21FMT=fullwebr&C21COM=S&S21CNR=10&S21P01=0&S21P02=0&S21P03=U=&S21STR=629.735.33%28062%29)

**Величенко Є. О.**

*Національний авіаційний університет,**Київ*

**ПОШУК нових знань з великої кількості інформації**

Згідно з деякими оцінками інформація подвоюється кожні 2-3 роки. Цунамі даних приходить з науки, бізнесу, інтернета та інших джерел. Пошук нових знань (далі Data Mining) включає в себе методи автоматичного аналізу, за допомогою яких приходиться практично добувати нові знання з великої кількості інформації.

Data Mining займається дослідженням і виявленням «машиною»(за доп. алгоритмів, засобами штучного інтелекту) в сирих даних приховані знання, які раніше не були відомі, нетривіальні, практично корисні та доступні для інтерпретації людиною.

Задачі Data Mining поділяють на:

* Задача класифікації. зводиться до визначення класу об'єкта за його характеристикам. При цьому безліч класів, до яких може бути віднесений об'єкт заздалегідь відомо.
* Задача регресії. Подібна задачі класифікації, однак значення параметра являється не кінечною множиною класів, а натуральним числом.
* При пошуку асоціативних правил метою є находження частих залежностей між представлених у вигляді правил і можуть бути використані як для кращого розуміння природи аналізуємих даних, так і для прогнозування появи подій.
* Задача кластеризації. Заключається в пошуку незалежних груп та їх характеристик у всій множині аналізуємих даних. Рішення цієї задачі допомагає краще зрозуміти дані. Крім цього, групування подібних об’єктів дозволяє зменшити їхню кількість, а відповідно і полегшити аналіз.

Досить великий інтерес виклиє сиквенціальний аналіз - виявлення закономірностей послідовності подій. Такий аналіз є різновидом задачі пошуку асоціативних правил. Він дає можливість з деякою долею ймовірності передбачити появу подій у майбутньому.

Вже існує досить багато готових програмних рішень для різних сфер. Однак, згідно різноманітним звітам (джерело: http://www.forbes.com/sites/louiscolumbus/

2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/) щорічний приріст витрат на розвиток систем аналітики та обробки даних становить 20% - 30%. Прогнозований сумарний обсяг витрат на системи обробки даних та аналітичні сервіси в 2015 році складе ~ 3млрд $. Отже поява нових програмних рішень Data Mining є досить перспективним напрямком.

*Науковий керівник – Є.Б.Артамонов, к.т.н*