**92.NYSK（-事件对股票的影响）**

1. 数据库网址

http://archive.ics.uci.edu/ml/datasets/NYSK

2. 数据库描述

【1.[数据集名称]数据集由[机构名或人名]采集；】The data used in our experiments were collected by E. Alpaydin, C. Kaynak, from Department of Computer Engineering,Bogazici University at July,1998.【2.用于[什么实验目的]】We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form.【3】

【4】The database has 5620 samples, respectively belong to optdigits.tra with 3823 samples and optidigits.tes with 1797 samples. The categories of network system include seven categories, as shown in Table 1.

Table 1 Category Distribution of Network System [根据数据库绘制]

|  |  |  |  |
| --- | --- | --- | --- |
| Invasion Categories | optdigits.tra | optdigits.tes | Total Number of Samples |
|  |  |  |  |
|  |  |  |  |
| Total number of samples in total |  |  | 10421 |

|  |  |
| --- | --- |
| **Abstract**: NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011). |  |

**Source:**

- AurÃ©lien Lauf (alu **'@'** amisw.com)   
- Leila Khouas (lkh **'@'** amisw.com)   
- Mohamed Dermouche (mde **'@'** amisw.com)

**Data Set Information:**

Documents are first obtained via a Web search using AMIEI: an integrated platform for delivering enterprise intelligence, developed by AMI Software ([[Web Link]](http://www.amisw.com/en)) with the following query: ``dsk'' OR ``strauss-kahn'' OR ``strauss-khan''.   
  
NYSK dataset was used to extract topic-sentiment correlation and evolution over time but may be used for other text mining tasks like topic extraction, sentiment analysis, etc.

**Attribute Information:**

Documents are then filtered and presented in XML format. All XML fields are self explanatory.