

Інтелектуальний аналіз даних сільськогосподарської галузі як елемента продовольчої безпеки України

Виконала:

Связінська Н.О., КА-43м

Науковий керівник:

д.т.н., проф. Бідюк П.І.

Об'єкт: структуровані та неструктуровані дані сільськогосподарської галузі

Предмет: моделі і методи інтелектуального аналізу даних

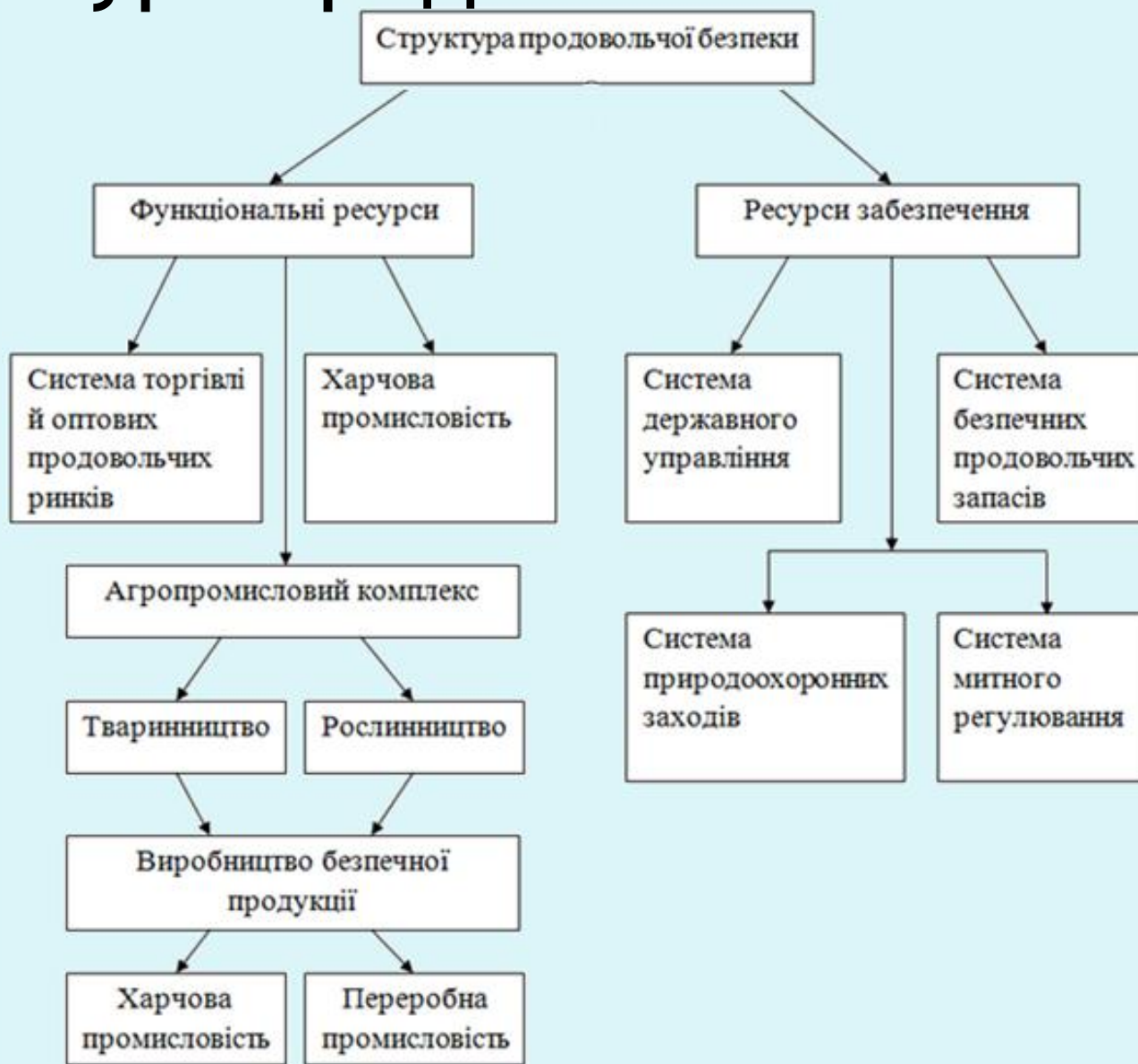
Завдання: застосування різних підходів та методологій, комплексний аналіз стану та тенденцій сільськогосподарської галузі з точки зору продовольчої безпеки.

Актуальність: про проблему продовольчої безпеки вперше заговорили у кінці минулого століття і з того часу вона не втрачає своєї актуальності. Сільськогосподарська ж галузь є надзвичайно важливою складовою продовольчої безпеки.

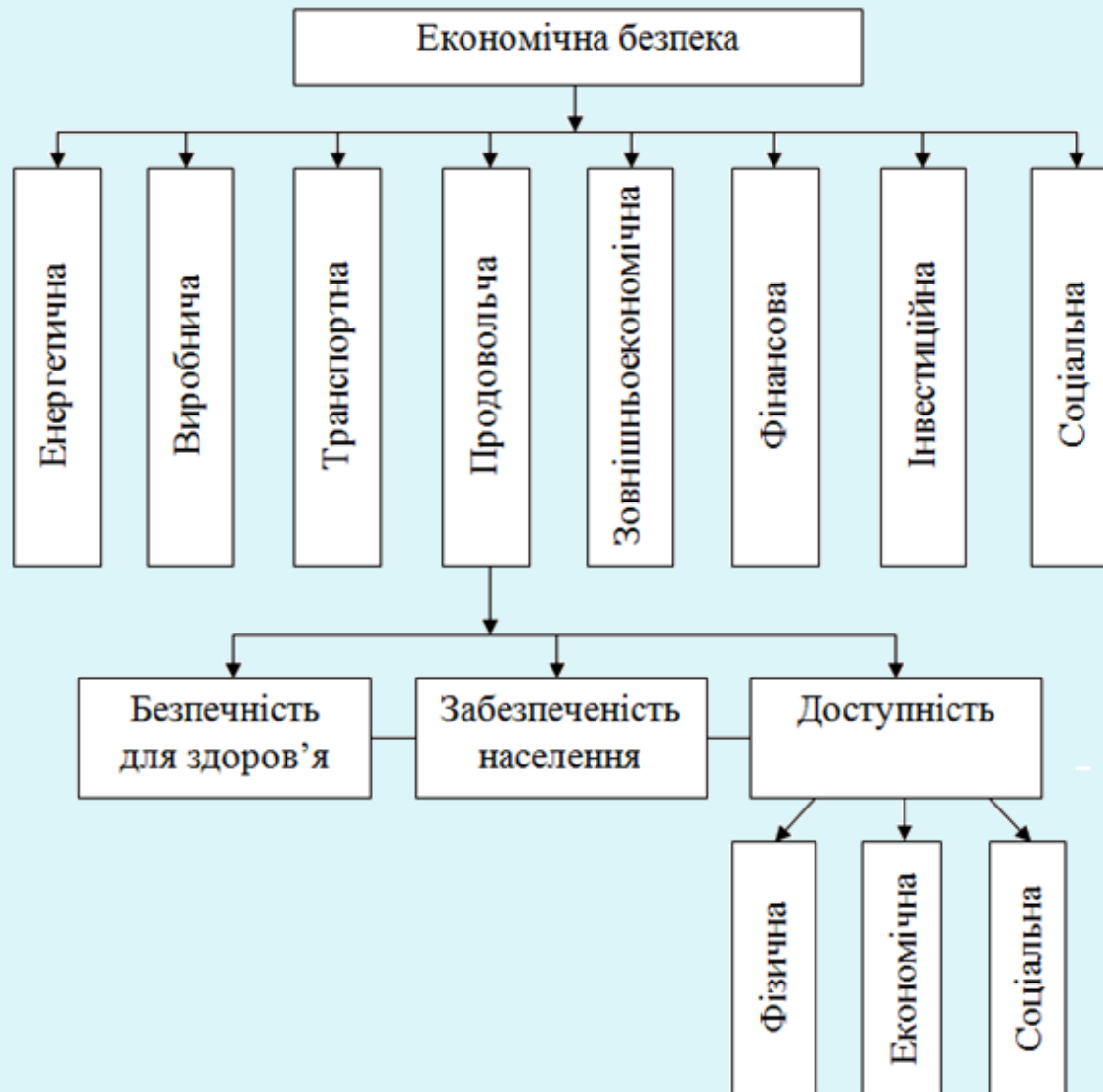
Інтелектуальний аналіз даних - дослідження та виявлення «машиною» (алгоритмами, засобами штучного інтелекту) в сирих даних прихованих, раніше невідомих, нетривіальних, практично корисних та доступних для інтерпретації людиною знань, необхідних для прийняття рішень у різних сферах людської діяльності.

Продовольча безпека - це стан, що характеризується постійним фізичним, соціальним та економічним доступом усіх людей до достатнього за обсягом, поживністю та безпечністю продовольства, яке необхідне для здорового та активного життя.

Структура продовольчої безпеки



Продовольча безпека у структурі економічної безпеки



Фактори рівня забезпеченості продовольчої безпеки

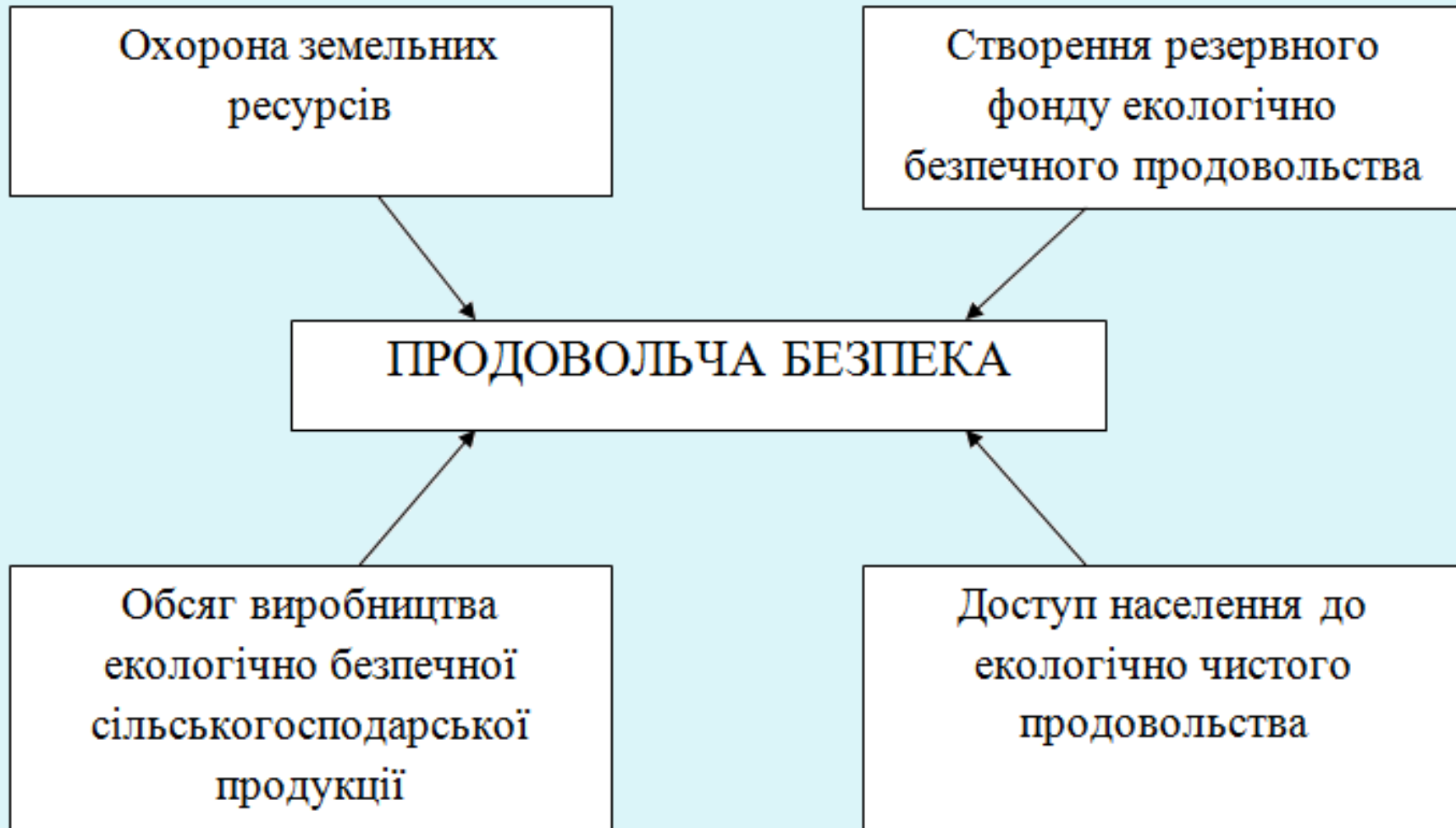


Схема процесу дослідження

Сбір даних

- Текстова аналітика
- Державна служба статистики
- Міжнародні профільні організації

Аналіз даних та математичне моделювання

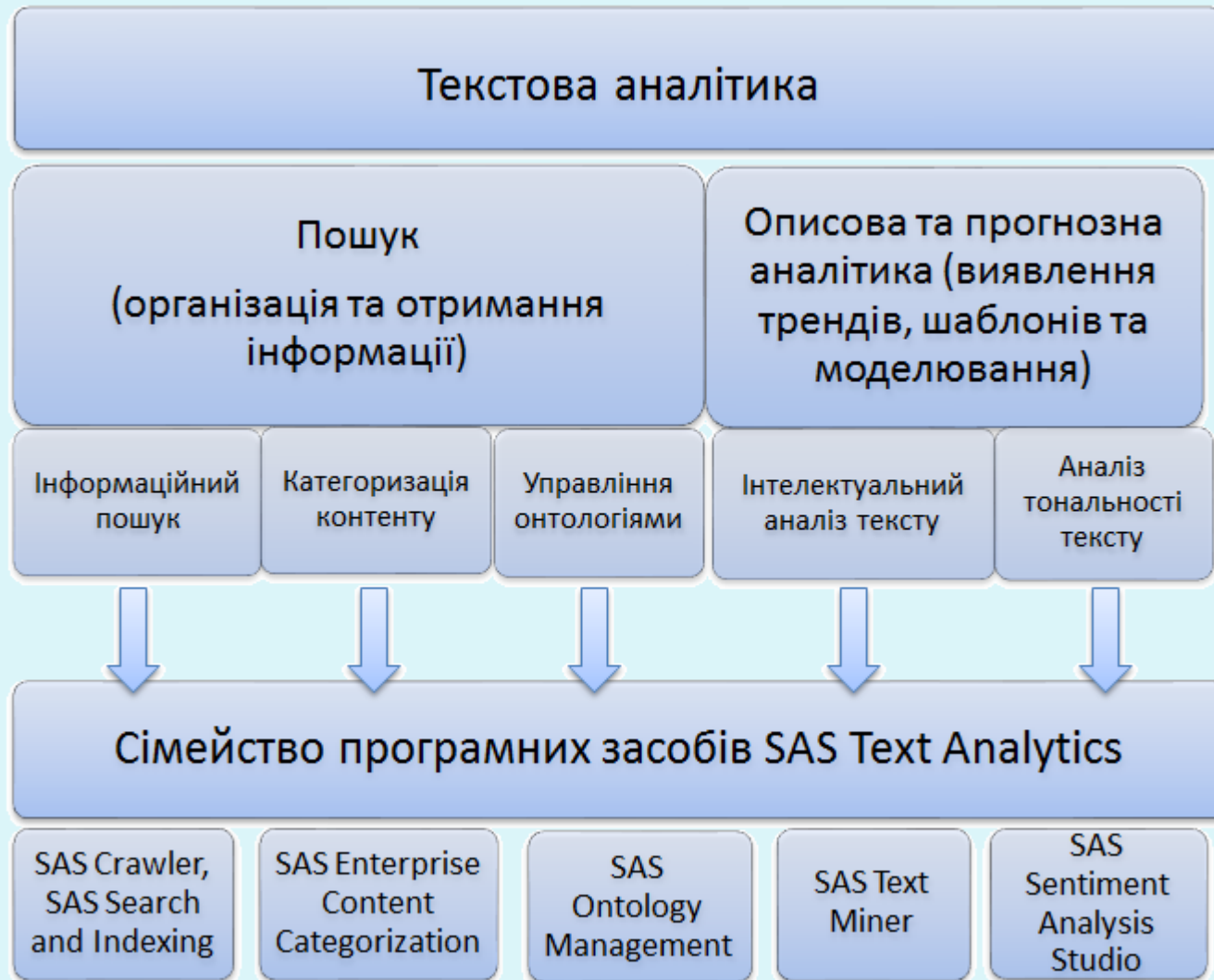
- Аналіз вхідних даних
- Трансформація даних
- Моделювання

Аналіз результатів

- Оцінка
- Формулювання висновків щодо стану продовольчої безпеки

Текстова аналітика

Методи текстової аналітики та відповідні програмні засоби SAS



Використані програмні засоби SAS

- SAS Information Retrieval Studio (SAS Crawler, SAS Markup Matcher)
- SAS Enterprise Content Categorization
- SAS Sentiment Analysis Studio



Джерела інформації

Назва Інтернет-ресурсу	Кількість завантажених статей
http://latifundist.com/blog	5649
http://agroua.net/news/	1324
http://obozrevatel.com/blogs	2741
http://blog.liga.net	11008
http://blogs.lb.ua	3538
http://agroportal.ua/views/blogs	1038
http://www.agro24.com.ua/blog	116
http://blogs.pravda.com.ua/	23180

Crawling sources

SAS Markup Matcher

File Edit Testing Help

Matcher

- Rules
 - URL
 - Title**
 - Date
 - Body
- Templates
- Namespace Mappings

Test Document - <http://ab.org.ua/import-dobriv-v-ukrayini-v-sichni-berezni-2016-viris-na-55/>

```


</a>
</noindex>
<noindex>
  <a href="http://connect.mail.ru/share?share_url=http://ab.org.ua/import-dobriv-v-ukrayini-v-sichni-berezni-2016-viris-na-55/"
  rel="nofollow" target="_blank">
    
  </a>
</noindex>
<noindex>
  <a href="http://www.livejournal.com/update.bml?event=http://ab.org.ua/import-dobriv-v-ukrayini-v-sichni-berezni-2016-viris-na-55/&subject=Імпорт добрив в Україні в січні-березні 2016 виріс на 55%" rel="nofollow" target="_blank"
  title="Опубликовать в своем блоге livejournal.com">
    
  </a>
</noindex>
</span>


<span class="ab-dostatti">
  <span class="ab-data_chas">

```

Test Results

Field	Value
Document	
Title	Імпорт добрив в Україні в січні-березні 2016 виріс на 55%
Body	За січень-березень 2016 року з України було експортовано добрив на загальну суму \$ 97,56 млн., Що на 42,9% м
URL	http://ab.org.ua/import-dobriv-v-ukrayini-v-sichni-berezni-2016-viris-na-55/
Date	http://ab.org.ua/2016/05/25/

Field Creation Rule

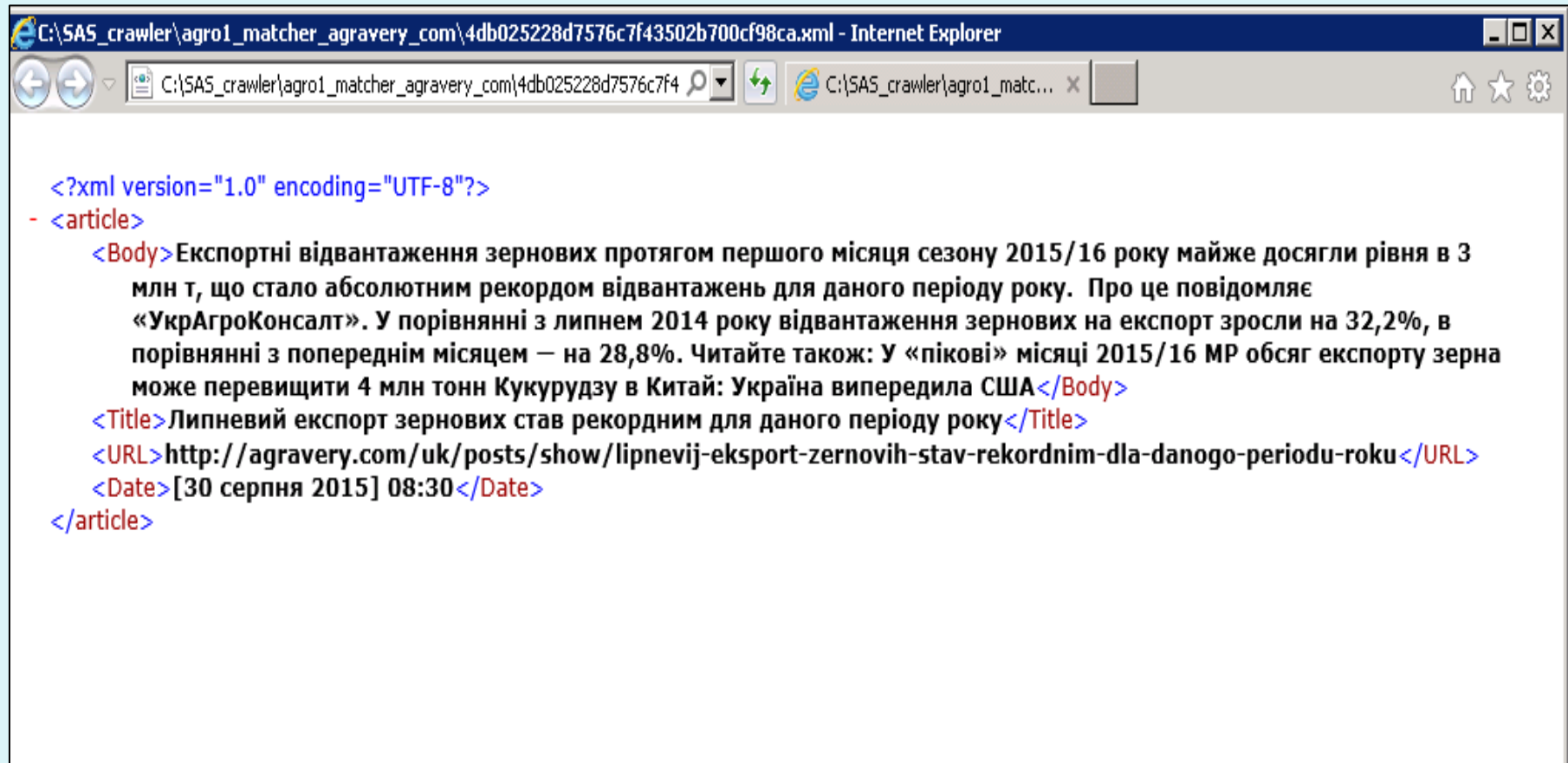
Field name:

Title

XPath expression:

```
/html/body/center/span/span[2]/span[2]/span[2]/h2/text()
```

Crawling sources



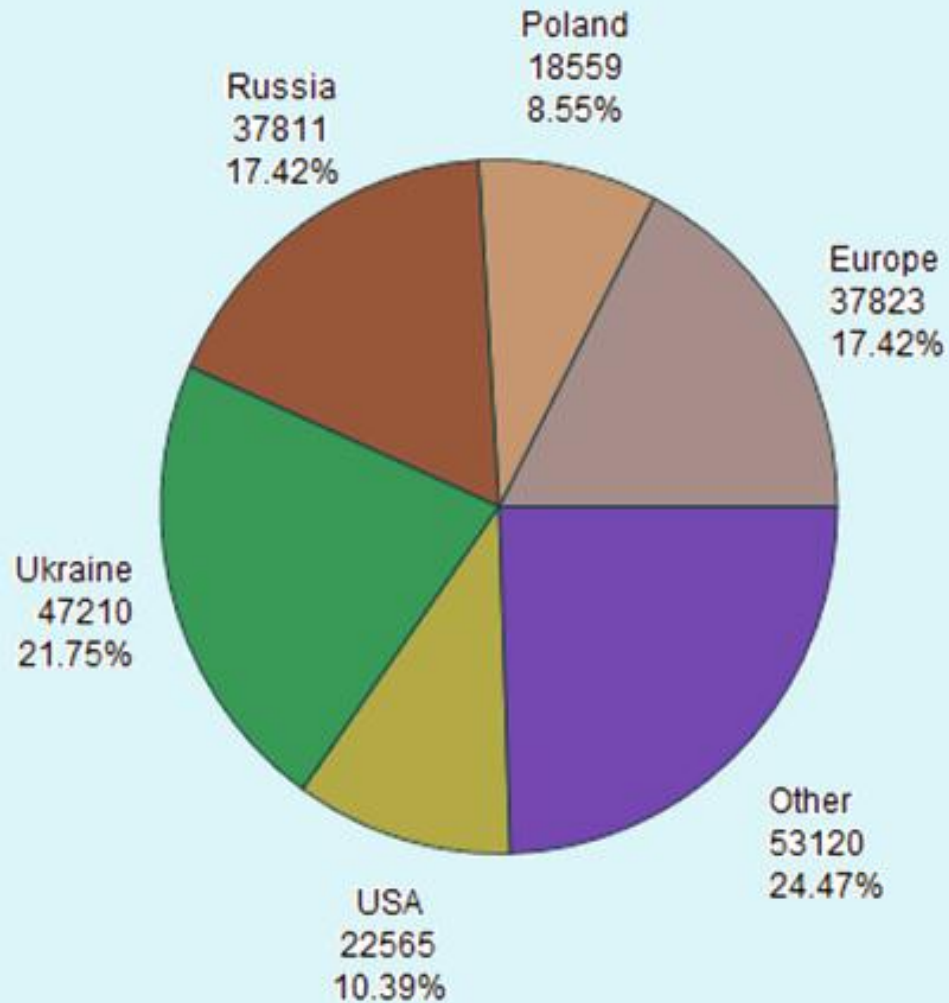
Analysis

The screenshot displays the SAS Content Categorization Studio interface. The main window shows a taxonomy tree on the left and a rule configuration panel on the right.

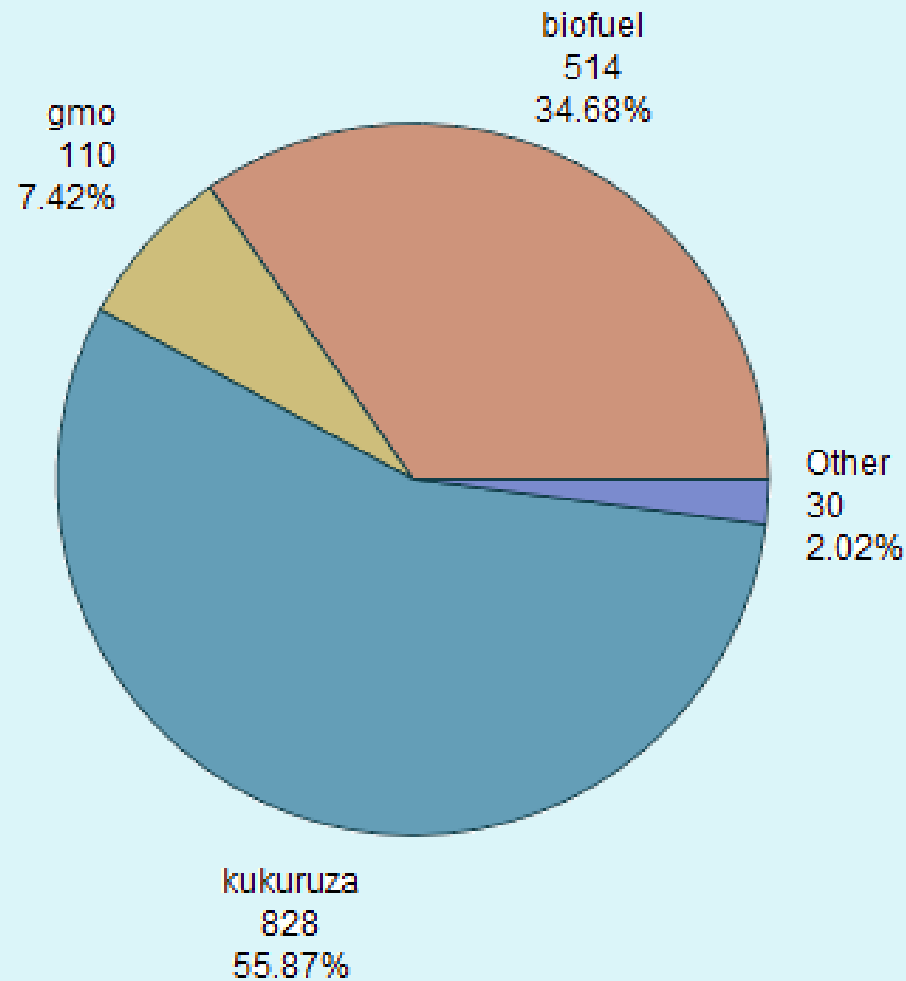
Left Panel (Taxonomy):

- kukuruza {kukuruza}
 - kukuruza_bury {kukuruza_bury}
 - kukuruza_buy {kukuruza_buy}
 - kukuruza_ling {kukuruza_ling}
 - kukuruza_production {kukuruza_production}
 - kukuruza_sell {kukuruza_sell}
 - kukuruza_storing {kukuruza_storing}
 - kukuruza_transportation {kukuruza_transportation}
 - kukuruza_export {kukuruza_export}
 - kukuruza_import {kukuruza_import}
- silos {silos}
 - silos_bury {silos_bury}
 - silos_buy {silos_buy}
 - silos_ling {silos_ling}
 - silos_production {silos_production}
 - silos_sell {silos_sell}
 - silos_storing {silos_storing}
 - silos_transportation {silos_transportation}
 - silos_export {silos_export}
 - silos_import {silos_import}
- sources_market {sources_market}
 - agroholding_market {agroholding_market}
 - agroholding_market_consumption {agroholding_market_consumption}
 - agroholding_market_consumption_up {agroholding_market_consumption_up}
 - agroholding_market_consumption_down {agroholding_market_consumption_down}
 - agroholding_market_consumption_high {agroholding_market_consumption_high}
 - agroholding_market_consumption_low {agroholding_market_consumption_low}

Країни, які найчастіше згадують



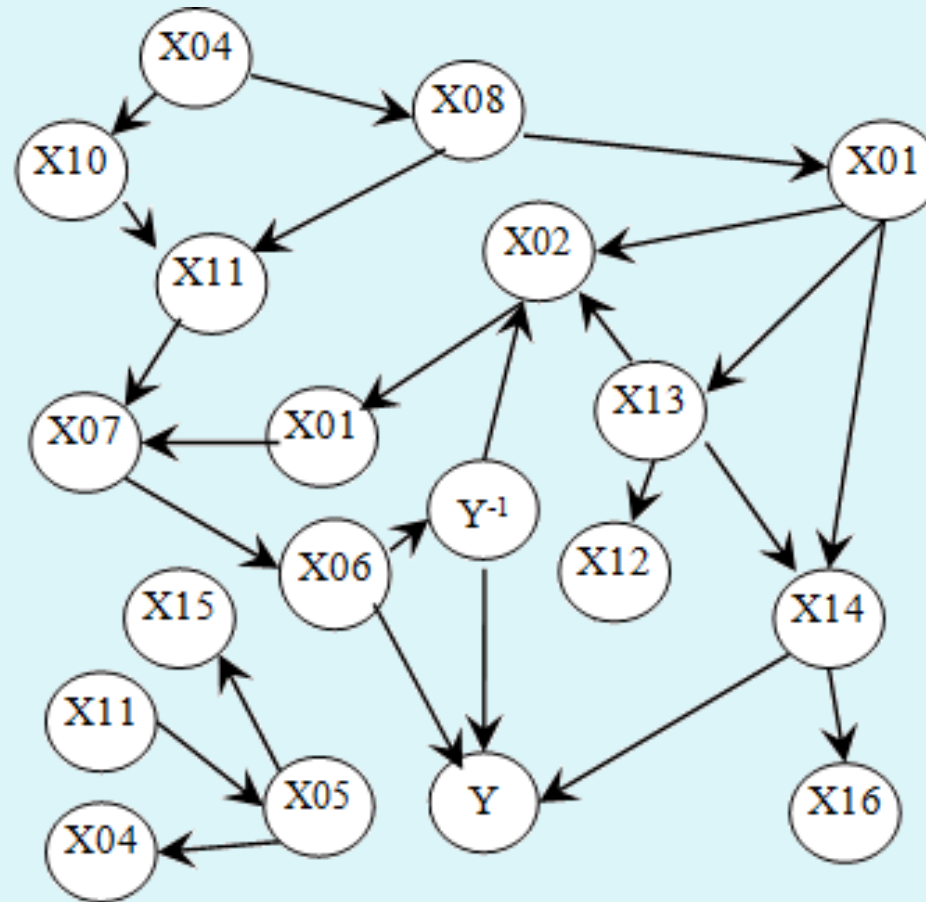
Найпопулярніші теми



Аналіз інвестиційної діяльності

Визначення ключових факторів

Мережа Байєса



Побудова регресійної моделі

$$Y(t) = 11429,9 + 2,7 * y(t-1) - 1,3 * x_{05}(t-1) - 3,9 * x_{06}(t-1) - 38,8 x_{14}(t-1),$$

де, x_{05} – демографічне навантаження, осіб;

x_{06} – оплата праці у сільському господарстві, грн;

x_{14} – коефіцієнт зносу основних засобів, %;

t – рік;

$Y(t)$ – капітальні інвестиції в агропромислове виробництво України, млн грн.

Розробка типології регіонів України

Використання RFM-аналізу

Класичний RFM-аналіз

$R(\text{Recency})=[1..5] \ \& \ F(\text{Frequency})=[1..5] \ \& \ M(\text{Monetary})=[1..5]$



$RFM=[111..555]$

Модифікований RFM-аналіз

$\text{Param_1}=[1..5] + \text{Param_2}=[1..5] + \dots + \text{Param_n}=[1..5]$



$RFM=[1..?]$

Використання RFM-аналізу

Група	Усереднений RFM- показник за 2010-2015 рр.	Регіон
Лідери	157-190	Вінницька обл., Дніпропетровська обл., Донецька обл., Київська обл., Кіровоградська обл., Полтавська обл., Харківська обл., Хмельницька обл., Черкаська обл.
Середняки	128-156	Запорізька обл., Івано-Франківська обл., Львівська обл., Миколаївська обл., Одеська обл., Сумська обл., Чернігівська обл., м. Київ
Аутсайders	76-127	Волинська обл., Житомирська обл., Закарпатська обл., Луганська обл., Рівненська обл., Тернопільська обл., Херсонська обл., Чернівецька обл.

Візуалізація типології регіонів України за допомогою системи SAS Enterprise Guide 7.1

- Побудова кластерної моделі
 - *Метод головних компонент*

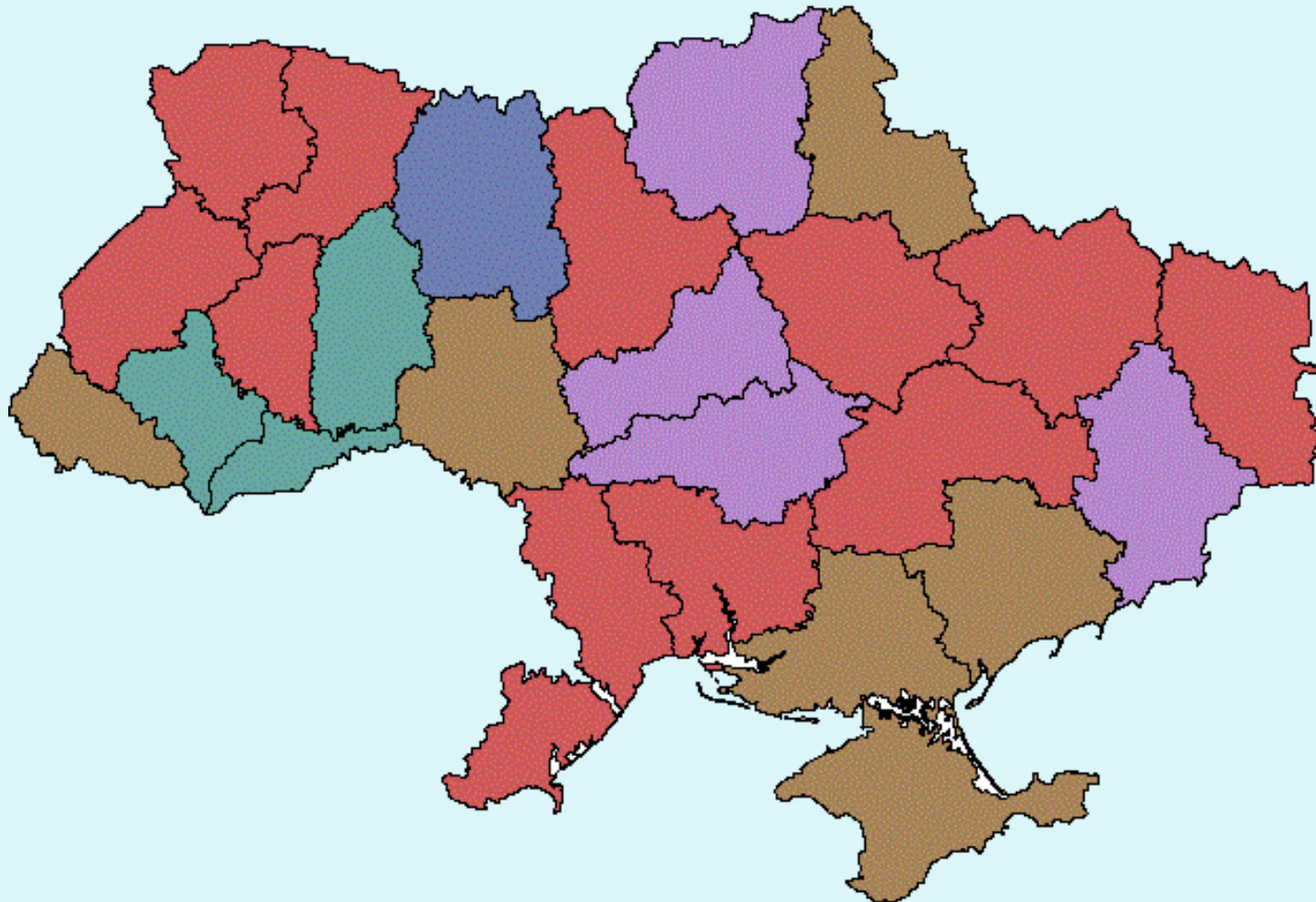
$$f_i = \sum_{j=1}^m U_{ij} X_j$$

- *Алгоритми кластеризації*
- Побудова інтерактивної карти

Групування регіонів за рівнем чутливості до притоку інвестицій

Номер кластеру	Регіон
1	Дніпропетровська обл., Донецька обл., Київська обл., Полтавська обл., Харківська обл., Черкаська обл.,
2	Вінницька обл., Кіровоградська обл., Хмельницька обл.
3	Запорізька обл., Львівська обл., АР Крим, Миколаївська обл., Одеська обл.,
4	Івано-Франківська обл., Сумська обл., Чернігівська обл.
5	Луганська обл., Херсонська обл.
6	Волинська обл., Житомирська обл., Закарпатська обл., Рівненська обл., Тернопільська обл., Чернівецька обл.

Типологія регіонів України



Висновки та перспективи дослідження

- *Проведено початковий аналіз галузі засобами текстової аналітики*
- *Побудовано регресійну модель для прогнозування обсягу інвестицій*
- *Розроблено типологію регіонів України за чутливістю до притоку інвестицій*

Матеріали, пов'язані з дослідженням

ISDMCI'2016

Тези конференції ISDMCI-2016 – *“Text Mining Analysis of Agriculture Internet Sources Using SAS Software”*

GISCTEEM-2016

Тези конференції GISCTEEM-2016 –
«Візуалізація типології регіонів України із використанням системи SAS Enterprise Guide 7.1»



Тези конференції SAIT-2016 – *«Застосування наївного та дереводоповненого байєсівських класифікаторів для прогнозування кредитоспроможності фізичних осіб»*



Наукова стаття у виданні «Наукові вісті НТУУ «КПІ»» - *«Моделювання фінансових ризиків з використанням наївного та доповненого деревом класифікаторів у формі байєсівських мереж»*

- *Авторське свідоцтво на програмний продукт*

Дякую за увагу