Honey origine Identification

- Identification for consumers
- Information for beekeepers on monofloral honey production
- Protected Labelisation
Rapport d'essai

Informations transmises
Vos références: 06/201
Lieu de production: Li

1. EXAMEN PHYSIQUE
   a) Essais accrédités
      - pH et acidité
      - Conductivité
      - Indice de sucre
      - HMFP (mg/kg)

2. Examen pollinique (non accrédité)
   - Analyse pollinique
   - Type de miel: Colza
   - Origine botanique: Colza

3. Examen organoleptique
   - Profil odorant et goût
   - Intensité des saveurs et sensations
   - Archétypes

Sugars
Pollens
Organoleptic Properties
Physicochemical parameters
Typical human analysis

physicochemical parameters
- Sugars
- Pollens
- Organoleptic Properties

GOAL OF BOOSTED DECISION TREES:
STATISTICAL DECISION SUPPORT
Decision Trees

Example of a 2-class Decision Tree

Separation of the space using the decision tree
Elements are classified in boxes depending on the variables and the nodes chosen
Decision Trees

Variable linked to each node:

\[ P = \frac{\text{nb}(A)}{\text{nb}(A) + \text{nb}(B)} \]

**P-factor**

\[ \text{Gini} = P(1 - P) \]

**Gini**

Gini factor measures the lack of purity on each subgroup (Giny = 0 for a perfect classification)

On each node \( \rightarrow \) Gini (left) + Gini (Right) is calculated

Program choose the variables and the level which minimizes sum of Gini
Decision Trees

Tree n°1:

Each partition of the space gets the name of the most represented population
Decision Trees

Function and errors associated to each partition:

Associated error:

\[ \text{error} = \frac{\text{nb misclassified individuals}}{\text{nb total individuals}} \]

**Quality index** \( \alpha \) of tree n°1 is linked to the associated error: \( \alpha \uparrow \) if error \( \downarrow \)

Before building a 2\(^{nd}\) tree, weights are associated to individuals:

\[ w_i = C \quad \text{if individual is well classified} \]
\[ w_i = Ce^{\alpha_i} \quad \text{if individual is misclassified} \]

Misclassified individuals get a higher weight.
Boosted Decision Trees

Boosting is an iteration of several trees

All the decision trees obtained are summed for a better classification and for a better stability of the answer.
Application to honey

- Database: analysis results of honey samples collected from 2009 to 2012 (4 years)
- Database was « cleaned »:
  - Monofloral honeys
  - Min. 3 individuals / class
  - Qualitative variables transformed in quantitative variables (1 – 0)
- Number of iterations: 50

Number of Individuals = 406
Number of botanical origins = 26
Number of variables = 316
Variable importance

- **Pollen** 48% (178 variables)
- **Sugar** 18% (15 variables)
- **Physicochemical** 15% (7 variables)
- **Organoleptic** 18% (114 variables)
- **Visual** 1% (2 variables)
Conductivity & pH: most discriminating physico-chemical parameters

Glucose, fructose, erlose & saccharose: Most discriminating sugars

Typical tastes

Color parameters
<table>
<thead>
<tr>
<th>Pollen Type</th>
<th>Dominant</th>
<th>Under-represented</th>
<th>Over-represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>P6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- Dominant pollens are those with high counts.
- Under-represented pollens are those with low counts.
- Over-represented pollens are those that are disproportionately high compared to others.
Outcomes for 2013 honeys

• Prediction realised on 75 honeys from 2013 with the Boosted Decision Tree (2009-2012)

• 13 monofloral honeys
  - 12 well classified (rapeseed, rosemary, palm)
  - 1 misclassified (jujube, class not yet in the database)
  → 100% good identification

• 62 polyfloral honeys
Outcomes for 2013 honeys

1) Monofloral honey

individu 14684  , class Colza , bien classe = TRUE

<table>
<thead>
<tr>
<th></th>
<th>Laboratory Analysis</th>
<th>Boosted Decision Tree Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLZA (RAPESEED)</td>
<td>Reference honey</td>
<td>COLZA (RAPESEED)</td>
</tr>
</tbody>
</table>
Application to honey

2) Polyfloral honey

<table>
<thead>
<tr>
<th>Laboratory Analysis</th>
<th>Boosted Decision Tree Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRUITIERS DOMINANT</td>
<td>FRUITIERS (FRUIT)</td>
</tr>
<tr>
<td>(FRUIT PREVAILING)</td>
<td>&amp;</td>
</tr>
<tr>
<td>&amp;</td>
<td>PISSENLIT (DANDELION)</td>
</tr>
<tr>
<td>SAULE (WILLOW)</td>
<td>&amp;</td>
</tr>
<tr>
<td></td>
<td>FRUITIERS &amp; SAULE</td>
</tr>
<tr>
<td></td>
<td>(FRUIT &amp; WILLOW)</td>
</tr>
</tbody>
</table>

individu 14672, class Fruitiers dominant, bien classe = FALSE

- type: Fruitiers, proba: 0.234889248178218
- type: Pissenlit, proba: 0.1977279975379
- type: Fruitiers & Saule, proba: 0.122222570157609
- type: Bourdaine, proba: 0.0824073914285493
- type: Colza, proba: 0.0818284402434684
- type: Acacia, proba: 0.080111747020175
- type: Saule, proba: 0.0792822741775859

High conductivity Dandelion pollens
Application to honey

individu 14653 , class Toutes fleurs , bien classe = FALSE

<table>
<thead>
<tr>
<th>Type</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colza</td>
<td>0.155861185724068</td>
</tr>
<tr>
<td>Agrumes</td>
<td>0.146318352854727</td>
</tr>
<tr>
<td>Acacia</td>
<td>0.138102209903404</td>
</tr>
<tr>
<td>Lavande</td>
<td>0.103182696553883</td>
</tr>
<tr>
<td>Trèfles</td>
<td>0.0814418172783824</td>
</tr>
<tr>
<td>Thym</td>
<td>0.0600130038971022</td>
</tr>
<tr>
<td>Sarrasin</td>
<td>0.0421238033479969</td>
</tr>
<tr>
<td>Pissenlit</td>
<td>0.0406941713287997</td>
</tr>
<tr>
<td>Fruitiers</td>
<td>0.0393152333730134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Laboratory Analysis</th>
<th>Boosted Decision Tree Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOUTES FLEURS (FLOWER HONEY) Romarin (Rosemary) &amp; Type Rosacées (Rosaceae) &amp; Fruitiers (Fruit)</td>
<td>COLZA (RAPESEED) &amp; AGRUMES (CITRUS) &amp; ROBINIER (ROBINIA) &amp; LAVANDE (LAVENDER)</td>
</tr>
</tbody>
</table>

No confidence on BDT Analysis
Conclusions

• Powerful tool

• Should be used in addition to human analysis

Boosted Decision Tree should be carried out after typical analysis → no influence
Pay attention to the results proposed

• Perspectives in the future:
  Enrichment of the database needed
  Model should be adapted to polyfloral honeys, prevailing types and honeydew