

PREDIVÍ: big data-based wine harvest prediction model

Summary

The project is based on the use of big data and machine learning technologies to develop prediction models that provide forecasts on quality parameters and harvest volumes in the wine industry. To that end, we used the information available to the companies and organisations participating in the project regarding production histories, plots, ripeness checks, sampling, etc. enabling the prediction to be made, together with the variables obtained in the campaign (meteorology, satellite images, etc.).

The model created greatly improved the efficiency (reduction of costs related to taking samples in the field) of the predictions of the two variables studied, with errors of around 10%.

The model has a powerful learning capability, which means the error will be reduced in future campaigns.

Objectives

The increasing variability of the volume and quality of wine production means that increasing investments are being made in resources and work by technical teams to obtain predictions for harvests. Technical teams currently use various systems (sampling, ripening checks, gauging, etc.) but the reliability of the results that these systems provide has great potential for improvement. The large number of variables affecting both quality and quantity of production (such as weather conditions, field characteristics, production areas, etc.) mean that obtaining reliable predictions using traditional approaches is a very complex task.

In order to reduce the error, the project will use big data technology which enables a large number of hyper-localised meteorological variables (both historical and predictive) to be combined with historical records for production and ripening checks, among other variables, thereby facilitating the creation of machine learning-based prediction models, to improve on the forecasts currently made by technical teams.

The main objective of the project is to provide organisations and actors in the wine industry with decision-making support tools to obtain advance information on harvest predictions, particularly regarding:

- 1- Production volume per plot
- 2- Production quality parameters (grade, acidity, pH, etc.)
- 3- Classification of the plot qualitative potential

Description of the actions carried out in the project

The following actions were carried out in order to achieve the objectives set out in the project:

A.- Evaluation and definition of variables and data sources for each model and creation of the database architecture.

B.- Creation and training of prediction models, test results and interaction with beneficiary companies in order to ensure the models are as reliable as possible.

D.- Validation of results and automation of the data collection and transmission system.

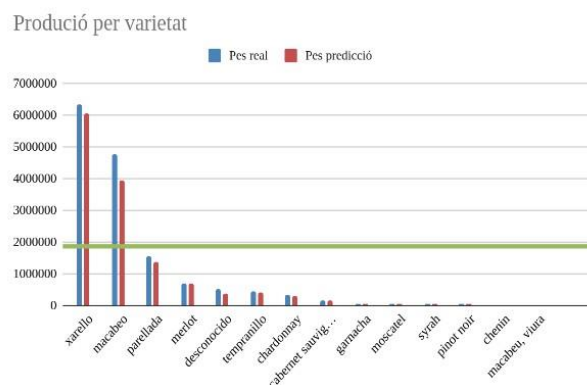
-
- ```

graph LR
 A((Problem Formulation)) --> B(Model Tuning)
 B --> C(Model Debugging)
 C --> D(Data Collection)
 D --> E(EDA)
 E --> F(Data Pre-processing)
 F --> G(Product ionization)
 F --> H(Feature Engineering)
 H --> I(Model Training)
 I --> J(Model Evaluation)
 J --> B

```

- ## Final results and practical recommendations

- The set of variables needed to create each prediction model were defined and validated. The final model uses: data from the fields currently in use, from the history of previous yields and harvests, from current and historical satellite images, and from current and historical hyperlocalised meteorological data.
- A prediction model based on the above variables was created, which gives results with errors of less than 10% in terms of quantity.



Predicted weight

- The technological development necessary to automate processes for data transmission between the current information systems of the beneficiary companies and the prediction models was carried out.
- A platform was created, which displayed the results of the prediction models based on graphs and maps representing geolocated variables.

## Conclusions

- The project has developed a model for predicting the quantity and quality of harvest with a margin of error similar to current methods, but which is much more efficient.
- If the historical data and those obtained automatically by the system (meteorological and satellite images) are used in conjunction with those from a single manual sampling (8 per field are possible in a conventional campaign) the margin of error is substantially reduced.
- The system engages in continuous learning, so the margin of error between the prediction and reality is expected to be increasing smaller in each campaign.
- The level of automation and the display systems for the results were considered optimal by the end users.

## Leader of the Operational Group

ORGANISATION: COVIDES, SCCL

CONTACT E-MAIL: [adv@covides.com](mailto:adv@covides.com)

## Coordinator of the Operational Group

ORGANISATION: INNOVI Association of Innovative Companies

CONTACT E-MAIL: [oficina@innovi.cat](mailto:oficina@innovi.cat)

## Other members of the Operational Group (grant recipients)

ORGANISATION: VITALPE, SAT

CONTACT E-MAIL: [advitalpe@vitalpesat.com](mailto:advitalpe@vitalpesat.com)

ORGANISATION: UNIÓ FRUITS, SCCL

## Other members of the Operational Group (not recipients of the grant)

ORGANISATION: Eurecat

ORGANISATION: Raw Data

## Subject area(s) of application

- |                                     |                                           |
|-------------------------------------|-------------------------------------------|
| <input checked="" type="checkbox"/> | Agricultural production system            |
| <input checked="" type="checkbox"/> | Agricultural practice                     |
| <input checked="" type="checkbox"/> | Agricultural equipment and machinery      |
| <input type="checkbox"/>            | Livestock farming and animal welfare      |
| <input checked="" type="checkbox"/> | Vegetable production and horticulture     |
| <input type="checkbox"/>            | Landscape / Territorial management        |
| <input type="checkbox"/>            | Pest and disease control                  |
| <input type="checkbox"/>            | Fertilisation and nutrient management     |
| <input type="checkbox"/>            | Soil management                           |
| <input type="checkbox"/>            | Genetic resources                         |
| <input type="checkbox"/>            | Forestry                                  |
| <input type="checkbox"/>            | Water management                          |
| <input type="checkbox"/>            | Climate and Climate Change                |
| <input checked="" type="checkbox"/> | Energy management                         |
| <input type="checkbox"/>            | Waste and by-product management           |
| <input type="checkbox"/>            | Biodiversity and environmental management |
| <input type="checkbox"/>            | Food quality/processing and nutrition     |
| <input type="checkbox"/>            | Supply chain, marketing and consumption   |

- ☐ Competitiveness and agricultural and forestry diversification
- ☐ General

## Geographical area(s) of application

### PROVINCE(S):

Barcelona, Tarragona, Lleida and Girona

### REGION(S):

All wine-producing regions

## Dissemination of the project: publications, seminars, multimedia, etc.

Both face-to-face and online seminars were held in order to maximise the impact of the project, including some events with agents in the sector.

- A face-to-face presentation of the project took place on 07/06/2019 at the 4th STRATEGIC IMMERSION, where Efoodprint and INCAVI presented the project to the members of INNOVI
- It was broadcast on Twitter to the general public on 19/12/2019. [LINK TO TWEET.](#)
- The results of the project were presented as part of a PATT seminar for winemakers and viticulturists on 03/11/2021. [LINK TO THE PROGRAMME.](#)
- A Latin American seminar took place on the VITIVIN platform. [LINK to the seminar](#)
- A news report was published on the INNOVI website in order to increase its visibility among the members. [LINK to the report.](#)

## Project website

<https://www.innovi.cat/predivi/>

## More information on the project

| PROJECT DATES            | TOTAL BUDGET              |
|--------------------------|---------------------------|
| Starting date: July 2019 | Total budget: €191,674.47 |
| End date: September 2021 | DARP funding: €78,333.37  |
| Current status: Executed | EU funding: €59,093.60    |
|                          | Own funding: €54,247.50   |

## With funding from:

Project funded through Operation 16.01.01 (Cooperation for Innovation) through the Catalan Rural Development Programme 2014-2020.

Order ARP/133/2017 of 21 June, approving the regulatory bases for grants for cooperation for innovation by promoting the creation of European Association for Innovation operational groups in the areas of agricultural productivity and sustainability and the execution of innovative pilot projects by those groups, and Resolution ARP/1282/2018, of 8 June, announcing the call for the grant.



Generalitat de Catalunya  
Departament d'Acció Climàtica,  
Alimentació i Agenda Rural



Fons Europeu Agrícola  
de Desenvolupament Rural:  
Europa inverteix en les zones rurals