# Real Bogus & Machine Learning for transient astronomy

Author:Bruno Sánchez
Mariano Domínguez
Marcelo Lares
Mario Díaz
Martín Beroiz

IATE - CONICET - Universidad Nacional de Córdoba (UNC)

17 de julio de 2015

Variability studies in astronomy refer to the fluctuation of brightness over time of an astronomical object.

We use variability studies for several things:

- Object classification
- NEO's studies
- Stellar systems ($\alpha$ Cen)
- Distance measures
- Cosmology and cosmography

There are several types of variability:

- Periodic Variability

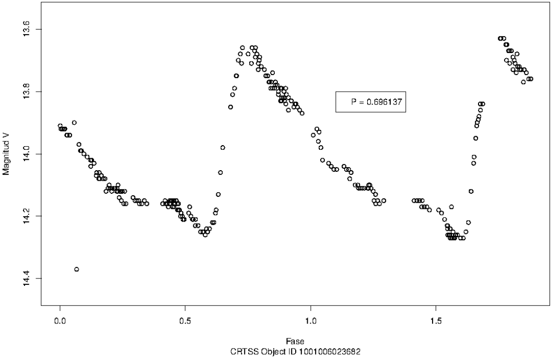



Figura: CRTSS light curve data for a RR Lyra star.

There are several types of variability:

- Periodic Variability
- Transient Variability

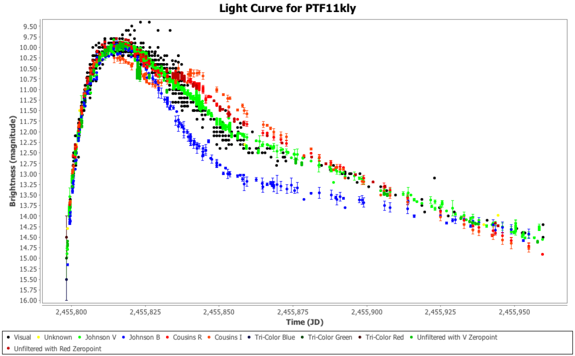



Figura: Supernova Type Ia lightcurve from PTF11kly. (PTF data).

There are several types of variability:

- Periodic Variability
- Transient Variability

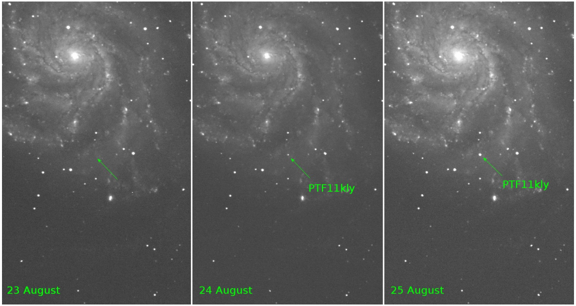


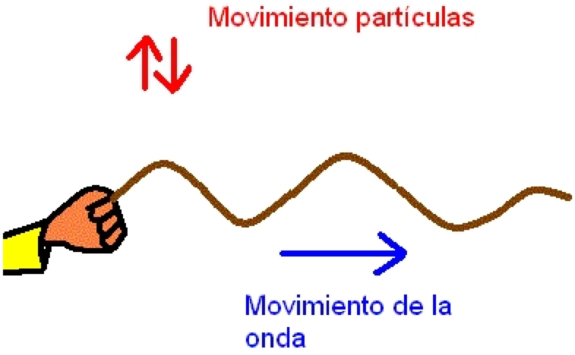

Figura: Supernova Type Ia (PTF11kly).

General Relativity (GR) admits on its equations the existence of **gravitational radiation**.

This Gravitational Waves (GW) would be **transversal waves**: induced perturbations arise in the perpendicular plane to the propagation direction.

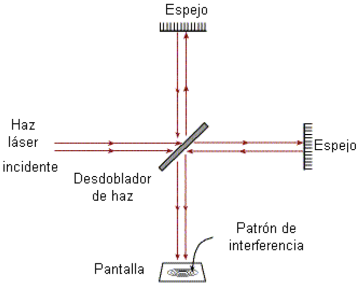
General Relativity (GR) admits on its equations the existence of **gravitational radiation**.

This Gravitational Waves (GW) would be **transversal waves**: induced perturbations arise in the perpendicular plane to the propagation direction.



To detect this ALIGO is using Michelson interferometers

To detect this ALIGO is using Michelson interferometers
For higher precision and sensibility this instruments need to be Km long



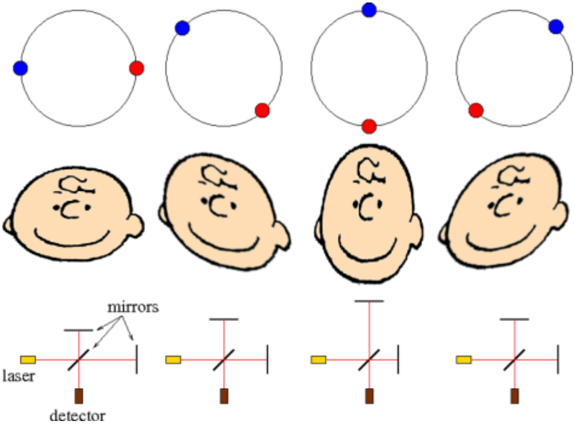




Figura: Perturbations induced by GW (from
http://www.einstein-online.info/spotlights/gravWav).

Localization errors of the candidates depends on N of detectors.

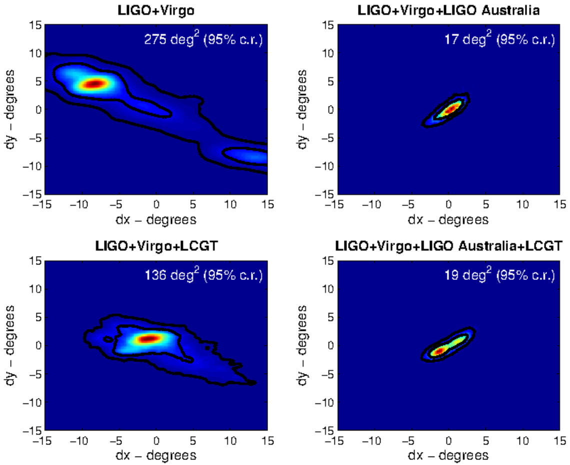



Figura: Sky localization of a low-limit SNR.

So ALIGO needs confirmation of candidates by a independent method.

So ALIGO needs confirmation of candidates by a independent method.
This independent tracer are the **Compact Object Mergers**.
This objects would produce bursts of EM radiation, (the so called "smoking gun").
In the optical range we find the Kilonova event.

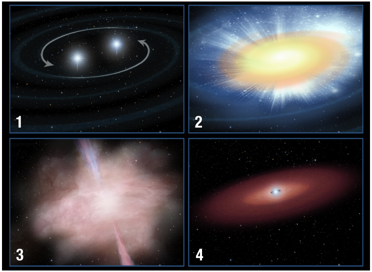



Figura: Compact object merger cartoon

There is only one observed Kilonova.

There is only one observed Kilonova.
So... what is needed to catch a Kilonova??

- Telescope time

There is only one observed Kilonova.
So... what is needed to catch a Kilonova??

- Telescope time
- Great number of observations

There is only one observed Kilonova.
So... what is needed to catch a Kilonova??

- Telescope time
- Great number of observations
- Be able to detect short timescale variability

There is only one observed Kilonova.
So... what is needed to catch a Kilonova??

- Telescope time
- Great number of observations
- Be able to detect short timescale variability
- High-speed data processing

There is only one observed Kilonova.
So... what is needed to catch a Kilonova??

- Telescope time
- Great number of observations
- Be able to detect short timescale variability
- High-speed data processing

There is only one observed Kilonova.
So... what is needed to catch a Kilonova??

- Telescope time
- Great number of observations
- Be able to detect short timescale variability
- High-speed data processing

Project TOROS - TORITOS


# Image analysis

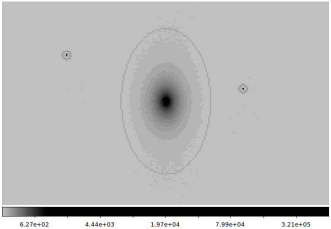
There are two different techniques:

- Aperture photometry

# Image analysis

There are two different techniques:

- Aperture photometry
  - Calculates the energy received on a given area of the image
  - $m_1 - m_2 = -2{,}5 log_{10} \left( \frac{E_1}{E_2} \right)$

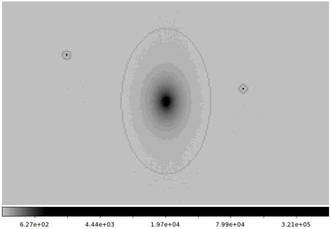



- Difference Image Analysis

# Image analysis

There are two different techniques:

- Aperture photometry
  - Calculates the energy received on a given area of the image
  - $m_1 - m_2 = -2{,}5 log_{10}\left(\frac{E_1}{E_2}\right)$



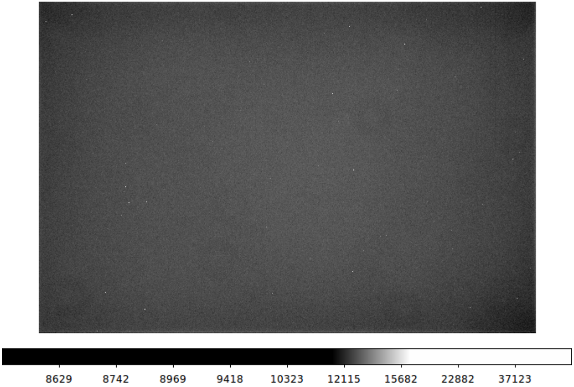

- Difference Image Analysis
  - It is like “fancy substracting” two images

# Image analysis

Our main approach relies on the second technique.

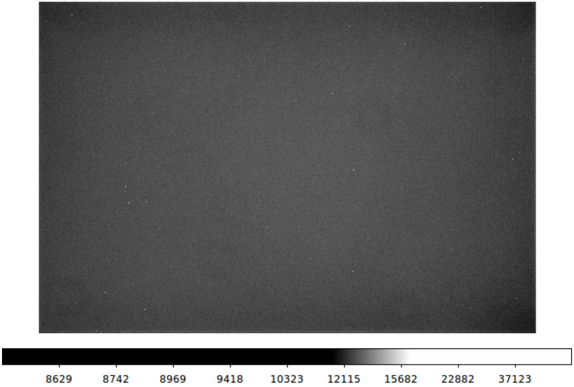
# Image analysis

Our main approach relies on the second technique.
In a perfect substraction we would be having a image like this:



# Image analysis

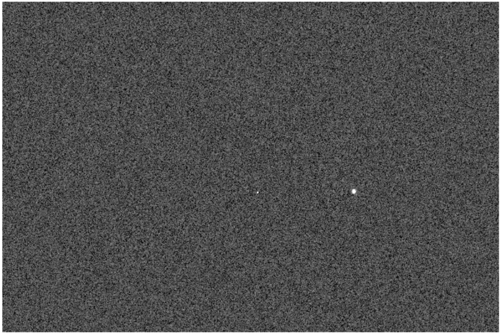
Our main approach relies on the second technique.
In a perfect substraction we would be having a image like this:




In this example there are not transients. Althoug we can see some random noise resembling stars.

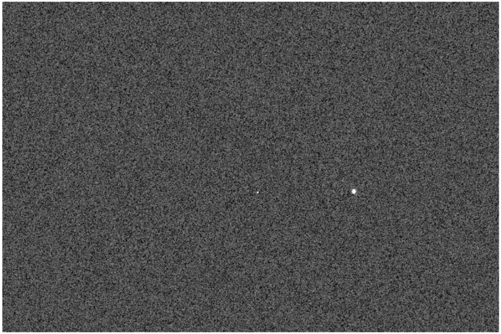
# Image analysis

This image presents transients:



# Image analysis

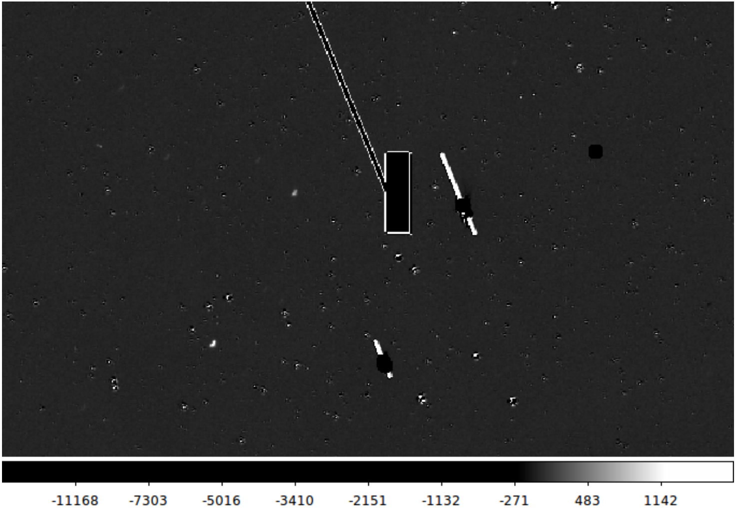
This image presents transients:



But this two are fake images.

# Image analysis

This is an actual subtraction from real data:



# Image analysis

In order to clean this kind of data we apply Statistical Learning methods, also called Machine Learning methods.

# Image analysis

In order to clean this kind of data we apply Statistical Learning methods, also called Machine Learning methods.

The classificator that separates false from real candidates is called Real-Bogus

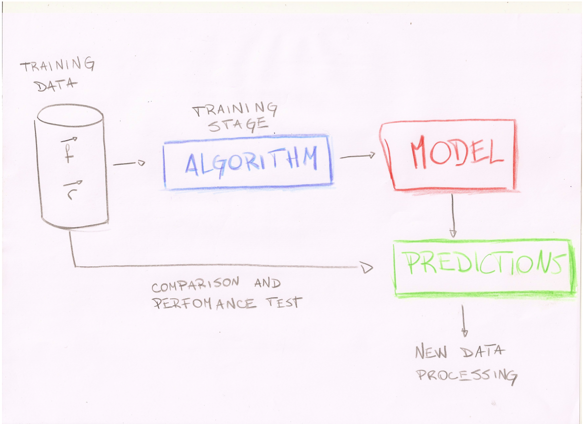
# Statistical Learning

First of all: *What is machine learning?*

# Statistical Learning

First of all: *What is machine learning?*

**Supervised learning**



# Statistical learning

The machine learning jargon includes different terms:

- **Objective class**: is the information we need to predict with a ML algorithm. It can be numeric, or categoric. The simplest is a binary objective class, such as R-B problem.

# Statistical learning

The machine learning jargon includes different terms:

- **Objective class**: is the information we need to predict with a ML algorithm. It can be numeric, or categoric. The simplest is a binary objective class, such as R-B problem.
- **Instances**: this are basically data points. Also called "examples", and they conform the datasets for training and testing stage.

# Statistical learning

The machine learning jargon includes different terms:

- **Objective class**: is the information we need to predict with a ML algorithm. It can be numeric, or categoric. The simplest is a binary objective class, such as R-B problem.
- **Instances**: this are basically data points. Also called “examples”, and they conform the datasets for training and testing stage.
- **Features**: are the measurables that we can extract from a given instance. Also represented as a vector $\vec{f_i}$ for the instance number $i$. These features can be also categoric or numeric.

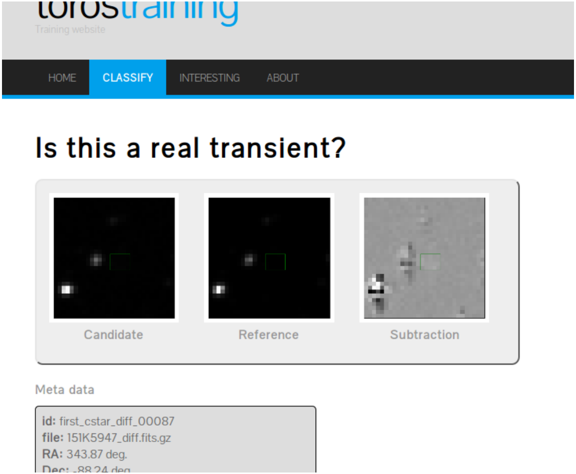
# Statistical learning

The machine learning jargon includes different terms:

- **Objective class**: is the information we need to predict with a ML algorithm. It can be numeric, or categoric. The simplest is a binary objective class, such as R-B problem.
- **Instances**: this are basically data points. Also called "examples", and they conform the datasets for training and testing stage.
- **Features**: are the measurables that we can extract from a given instance. Also represented as a vector $\vec{f_i}$ for the instance number $i$. These features can be also categoric or numeric.
- **Labeled data**: these data are instances that we are going to use as training set. Labeled means that you already know which is the true class for each instance.

# Statistical Learning

So a dataset with a correct labeling of different objects is needed to obtain a classificator.
This can be achieved for example by using citizen science: `http://toros-dev.no-ip.org`



# Algorithms implemented

Implemented three algorithms from ML:

- **Naive Bayes**
- **Logistic Regression**
- **Random Forest**

# Algorithms implemented

Implemented three algorithms from ML:

- **Naive Bayes**
- **Logistic Regression**
- **Random Forest**

# Algorithms implemented

Implemented three algorithms from ML:

- **Naive Bayes** Uses the Bayes's theorem in order to estimate probabilities of belonging to different classes.

$$P(c = R|\vec{f_i}) = \frac{P(\vec{f_i}|c = R)P(c = R)}{P(\vec{f_i})}$$

- **Logistic Regression**
- **Random Forest**

# Algorithms implemented

Implemented three algorithms from ML:

- **Naive Bayes**
- **Logistic Regression** This algorithm assumes that probabilities of belonging to a given class can be linearly interpolated after a simple transformation:

$$logi(p_R) = \ln(\frac{p_R}{1 - p_R}) = \alpha + \beta \times \vec{f_i}$$

- **Random Forest**

# Algorithms implemented

Implemented three algorithms from ML:

- **Naive Bayes**
- **Logistic Regression**
- **Random Forest** This algorithm works growing decision trees by using two parameters: $N_{tree}$ the number of decision trees to be trained, and $N_f$ the number of random features on each tree. After training the decision is made taking into account all the trees votes.

# Constructing traning dataset

We simulated data giving a balanced set of 9202 *bogus* y 9120 *reals*.

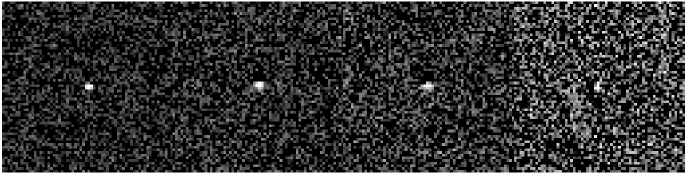


Figura: Stamps of bogus simulated objects.

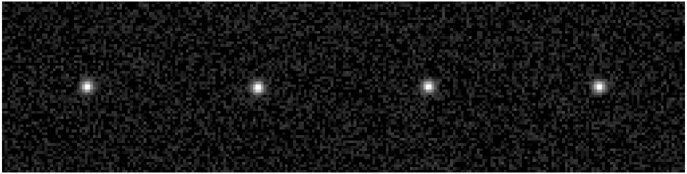


Figura: Simulated real objects.

# Procedure

- We implemented a **vectorizer**, which simply uses the image pixel data and extracts a huge set of features -it can extract almost 3000 features-, and we used a reduced **set of 1028 features**.

# Procedure

- We implemented a **vectorizer**, which simply uses the image pixel data and extracts a huge set of features -it can extract almost 3000 features-, and we used a reduced **set of 1028 features**.
- After that we implemented **WEKA Explorer suit** to analyze the data. It contains several standard ML tools, including **feature selection**.

# Procedure

- We implemented a **vectorizer**, which simply uses the image pixel data and extracts a huge set of features -it can extract almost 3000 features-, and we used a reduced **set of 1028 features**.
- After that we implemented **WEKA Explorer suit** to analyze the data. It contains several standard ML tools, including **feature selection**.
- From this tools we employed two feature selection routines: PCA, and CfSubsetEval from Hall (1998). This two algorithms reduce the 1028 set of features in order to discard interdependencies.

# Procedure

- We implemented a **vectorizer**, which simply uses the image pixel data and extracts a huge set of features -it can extract almost 3000 features-, and we used a reduced **set of 1028 features**.
- After that we implemented **WEKA Explorer suit** to analyze the data. It contains several standard ML tools, including **feature selection**.
- From this tools we employed two feature selection routines: PCA, and CfSubsetEval from Hall (1998). This two algorithms reduce the 1028 set of features in order to discard interdependencies.
- PCA threw 310 principal components, giving also the transformation matrix, and features vectors already projected into the PCA space.

# Procedure

- We implemented a **vectorizer**, which simply uses the image pixel data and extracts a huge set of features -it can extract almost 3000 features-, and we used a reduced **set of 1028 features**.
- After that we implemented **WEKA Explorer suit** to analyze the data. It contains several standard ML tools, including **feature selection**.
- From this tools we employed two feature selection routines: PCA, and CfSubsetEval from Hall (1998). This two algorithms reduce the 1028 set of features in order to discard interdependencies.
- PCA threw 310 principal components, giving also the transformation matrix, and features vectors already projected into the PCA space.
- The second method found that the best features were 39, and so we used this subsample to train the classificators.

# FOM

We call Figure of Merit to a set of estimators of a certain test perfomance. During every decistion process we deal with two possible errors:

| - | Desition | |
|---|---|---|
| Actual State | Reject $H_0$ | No reject $H_0$ |
| $H_0$ True | Type I Error | Correct |
| $H_0$ False | Correct | Type II Error |

This also represents the so called "Confusion Matrix"

# FOM

Some common quantities that measure the perfomance of a test are FDR, TPR, y FPR.

- FDR *False Discovery Rate* is the probability of making a Type I mistake, given that you already rejected $H_0$
- TPR *True Positive Rate*) is the probability of rejecting $H_0$ given that it is false, or 1- the probability of making a Type I mistake
- FPR (o *False Positive Rate*) is the probability of making a Type II mistake

# FOM

There is a compromise between FPR and TPR.
This compromise serves the preferences of the test's user for the confidence This curve is called *Receiver Operating Characteristic* or ROC.
The most used estimator is called AUC for *Area under the Curve*

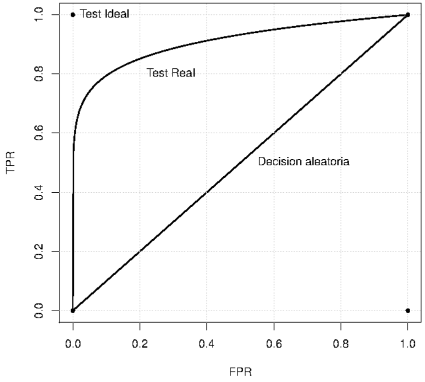



Figura: Curva ROC

# FOM

There are others FoM for different purpouses. FPR and TPR can be "fooled" in the presence of an unbalanced training set.
$N_B \sim \epsilon \times N_R$
In some data sets $\epsilon$ goes from 30 to 100. So the training set is unbalanced.

$$Precision = P(Y = 1|\hat{Y} = 1)$$
$$Recall = P(\hat{Y} = 1|Y = 1)$$
$$Specificity = P(\hat{Y} = 0|Y = 0)$$

Recall and Specificity relie on conditionals from the **true class labels**.
Precision relies on conditionals from **your estimate** of the true class label.

# FOM

**The Relationship Between Precision-Recall and ROC Curves**

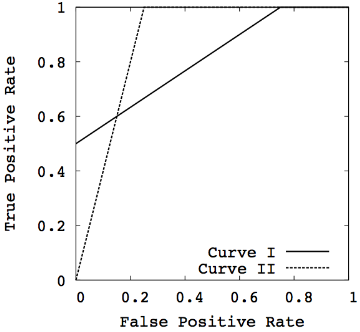



*Figure 11.* Comparing AUC-ROC for Two Algorithms

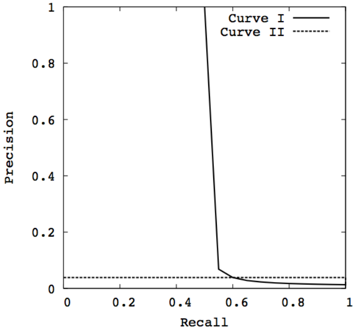
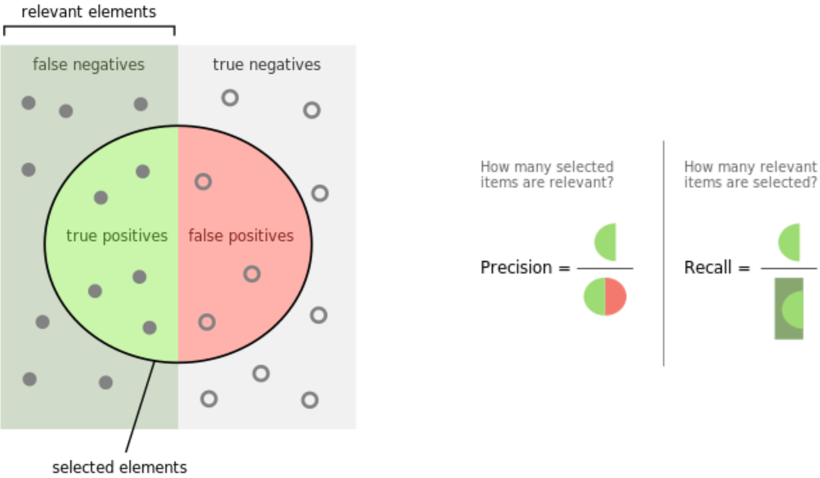


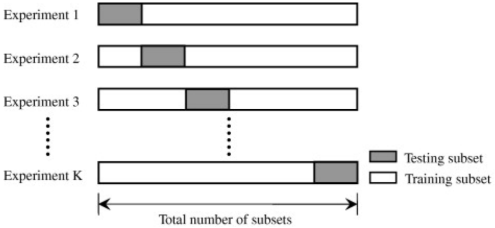

*Figure 12.* Comparing AUC-PR for Two Algorithms

# FOM


relevant elements
false negatives
true negatives
true positives
false positives
selected elements
How many selected items are relevant?
Precision =
How many relevant items are selected?
Recall =

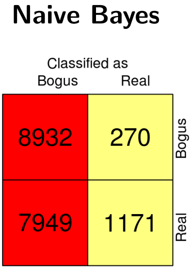
# K-fold cross validation

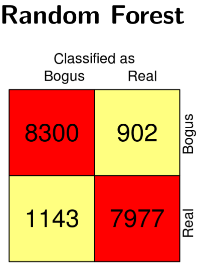
The perfomance was calculated using K-Fold Cross Validation algorithm, which divides the dataset into **K slices randomly sampled** and trains in K-1 folds while testing over the fold unused.

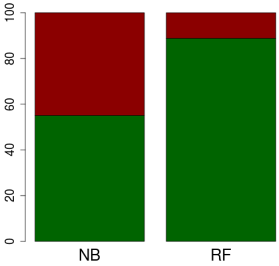



Since this can be performed K times, the testing gives K measures, and then those are combined to extract the TPR and FPR statistics.

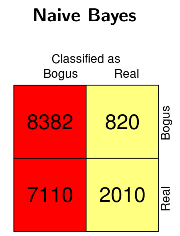
# Complete sample results

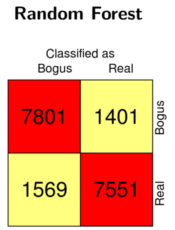


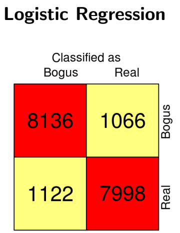



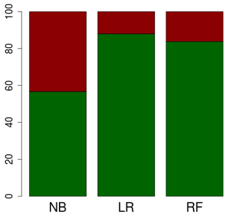



# PCA transformed sample results



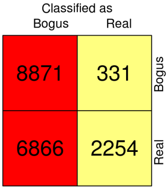



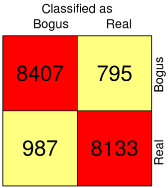



# Cfs selected features results

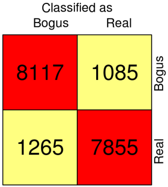
**Naive Bayes**

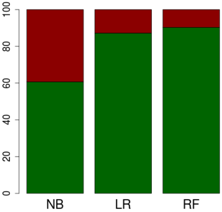



**Random Forest**




**Logistic Regression**



# Summary and conclusions

- We implemented Machine Learning algorithms for classification in astronomical images, using a vectorizer and feature selection methods.

# Summary and conclusions

- We implemented Machine Learning algorithms for classification in astronomical images, using a vectorizer and feature selection methods.
- The results showed a powerful tool that can be employed easily and with high confidence.

# Summary and conclusions

- We implemented Machine Learning algorithms for classification in astronomical images, using a vectorizer and feature selection methods.
- The results showed a powerful tool that can be employed easily and with high confidence.
- The flexibility of these methods make them exceptionally expandable to other situations.

# Future work

- Use this for TOROS/TORITOS (deployment on spring).
- Classify light curves

# Future work

- Use this for TOROS/TORITOS (deployment on spring).
- Classify light curves

**PREGUNTAS?? :D**