# Feature Subset Selection by Bayesian network-based optimization

Inza, I Larrañaga, P Etxeberria, R Sierra, B

A new method for Feature Subset Selection in machine learning, FSS-EBNA (Feature Subset Selection by Estimation of Bayesian Network Algorithm), is presented. FSS-EBNA is an evolutionary, population-based, randomized search algorithm, and it can be executed when domain knowledge is not available. A wrapper approach, over Naive-Bayes and ID3 learning algorithms, is used to evaluate the goodness of each visited solution. FSS-EBNA, based on the EDA (Estimation of Distribution Algorithm) paradigm, avoids the use of crossover and mutation operators to evolve the populations, in contrast to Genetic Algorithms. In absence of these operators, the evolution is guaranteed by the factorization of the probability distribution of the best solutions found in a generation of the search. This factorization is carried out by means of Bayesian networks. Promising results are achieved in a variety of tasks where domain knowledge is not available. The paper explains the main ideas of Feature Subset Selection, Estimation of Distribution Algorithm and Bayesian networks, presenting related work about each concept. A study about the 'overfitting' problem in the Feature Subset Selection process is carried out, obtaining a basis to define the stopping criteria of the new algorithm. (C) 2000 Elsevier Science B.V. All rights reserved.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000165320300005&DestLinkType=FullRecord&DestApp=WOS_CPL)

# Linkage Problem, Distribution Estimation, and Bayesian Networks

Pelikan, Martin Goldberg, David E. Cantu-Paz, Erick

This paper proposes an algorithm that uses an estimation of the joint distribution of promising solutions in order to generate new candidate solutions. The algorithm is settled into the context of genetic and evolutionary computation and the algorithms based on the estimation of distributions. The proposed algorithm is called the Bayesian Optimization Algorithm (BOA). To estimate the distribution of promising solutions, the techniques for modeling multivariate data by Bayesian networks are used. The BOA identifies, reproduces, and mixes building blocks up to a specified order. It is independent of the ordering of the variables in strings representing the solutions. Moreover, prior information about the problem can be incorporated into the algorithm, but it is not essential. First experiments were done with additively decomposable problems with both nonoverlapping as well as overlapping building blocks. The proposed algorithm is able to solve all but one of the tested problems in linear or close to linear time with respect to the problem size. Except for the maximal order of interactions to be covered, the algorithm does not use any prior knowledge about the problem. The BOA represents a step toward alleviating the problem of identifying and mixing building blocks correctly to obtain good solutions for problems with very limited domain information.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000208114700004&DestLinkType=FullRecord&DestApp=WOS_CPL)

# Feature subset selction using probabilistic tree structures.: A case study in the survival of cirrhotic patients treated with TIPS

Inza, I Merino, M Larrañaga, P Quiroga, J Sierra, B Girala, M

The transjugular intrahepatic portosystemic shunt (TIPS) is an interventional treatment for cirrhotic patients with portal hypertension. In the light of our medical staff's experience, the consequences of the TIPS are not homogeneous for all the patients and a subgroup of them dies in the first six months after the TIPS placement. Actually, there is no risk indicator to identify this group, before treatment. An investigation for predicting the survival of cirrhotic patients treated with TIPS is carried out using a clinical database with 107 cases and 77 attributes. Naive-Bayes, C4.5 and CN2 supervised classifiers axe applied to identify this group, The application of several Feature Subset Selection (FSS) techniques has significantly improved the predictive accuracy of these classifiers and considerably reduced the amount of attributes in the classification models. Among FSS techniques, FSS-TREE, a new randomized algorithm inspired on the EDA (Estimation of Distribution Algorithm) paradigm, has obtained the best accuracy results.

[Paper](https://www.webofscience.com/api/gateway?GWVersion=2&SrcApp=edas&SrcAuth=WosAPI&KeyUT=WOS:000171225400014&DestLinkType=FullRecord&DestApp=WOS_CPL)