

Feature Selection

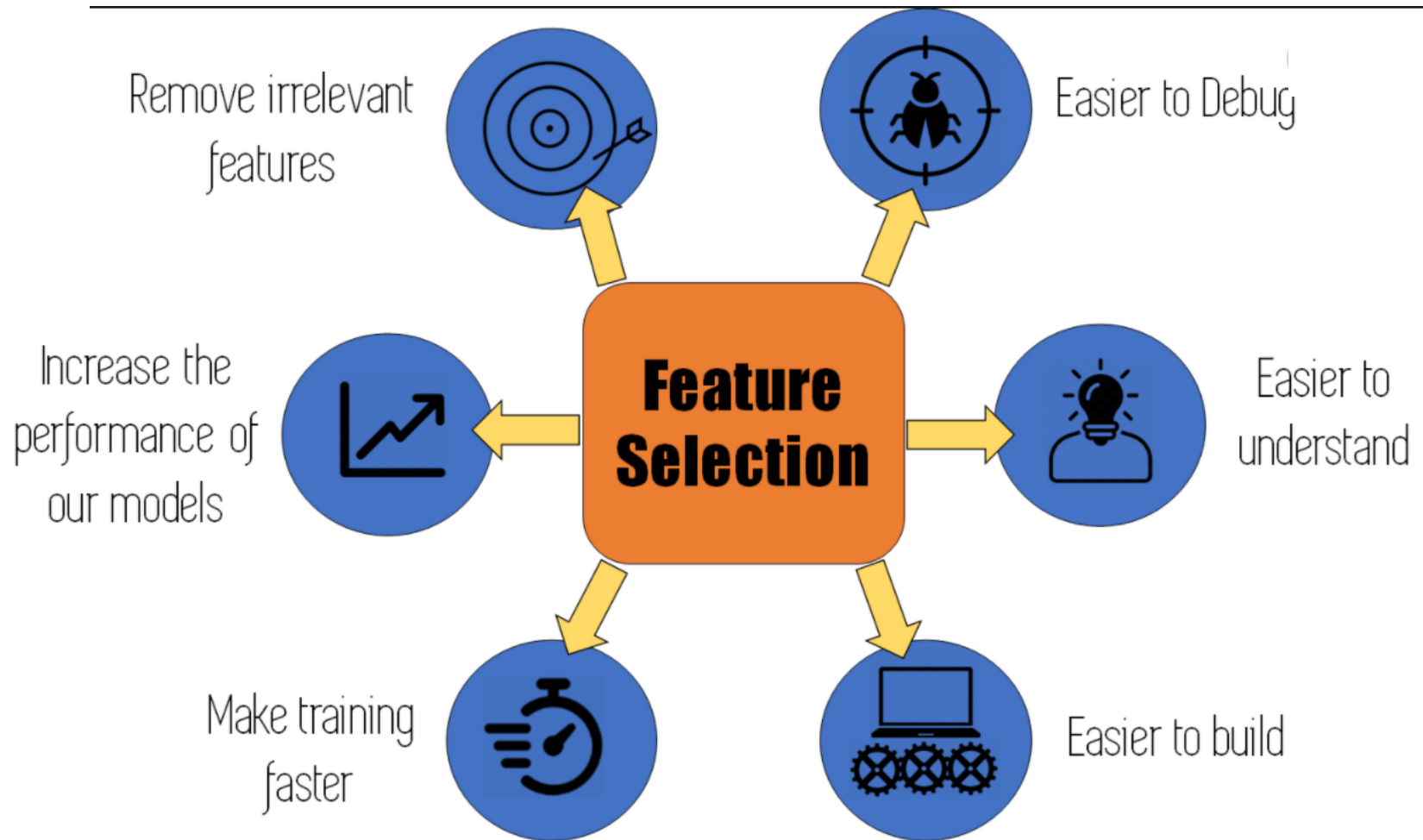
CS5056 Data Analytics

Francisco J. Cantú, Héctor Ceballos

Tecnológico de Monterrey

March 3, 2020

February-June, 2021



Feature Selection

- **Feature Selection**
- Feature selection is a process where you automatically select those features in your data that contribute most to the prediction variable or output in which you are interested.
- Having irrelevant features in your data can decrease the accuracy of many models, especially linear algorithms like linear and logistic regression.
- Three benefits of performing feature selection before modeling your data are:
- **Reduces Overfitting:** Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy:** Less misleading data means modeling accuracy improves.
- **Reduces Training Time:** Less data means that algorithms train faster.



Feature Selection Criteria



- Correlation
- p-values
- Accuracy
- Entropy
- Information gain
- Variance
- Parameter weights
- Scores

Feature Selection with sklearn and Pandas

Feature selection can be done in multiple ways but there are broadly 3 categories of it:

1. Filter Method
2. Wrapper Method
3. Embedded Method



Feature Selection

Do the Hands-on Exercises in:

<https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b>

Other Feature Selection Algorithms

- The classes in the [sklearn.feature_selection](https://scikit-learn.org/stable/modules/feature_selection.html) module can be used for feature selection/dimensionality reduction on sample sets, either to improve estimators' accuracy scores or to boost their performance on very high-dimensional datasets.

Francisco J. Cantú-Ortiz, PhD

Professor of Computer Science and Artificial Intelligence
Tecnológico de Monterrey
Enago-Academy Advisor for Strategic Alliances

E-mail: fcantu@tec.mx, fjcantor@gmail.com

Cel: +52 81 1050 8294, SNI-2 CVU: 9804

Personal Page: <http://semtech.mty.itesm.mx/fcantu/>

Facebook: [fcantu](#); Twitter: [@fjcantor](#); Skype: [fjcantor](#)

Orcid: 0000-0002-2015-0562

Scopus ID:6701563520

Researcher ID: B-8457-2009

https://www.researchgate.net/profile/Francisco_Cantu-Ortiz

<https://scholar.google.com.mx/citations?hl=es&user=45-uuK4AAAAJ>

<https://itesm.academia.edu/FranciscoJavierCantuOrtiz>

Ave. Eugenio Garza Sada No. 2501, Monterrey N.L., C.P. 64849, México