



IPS Paper

Applied Machine Learning for Central Bank Statistics: Supervised Models for The Detection Of Subsidized Housing Complexes In Chile

Author: Mr Martin Rebolledo

Coauthors: Martin Rebolledo

Submission ID: 1778

Reference Number: 1778

Presentation File

[abstracts/ottawa-2023_488c130caaadd379dbf45084b6ecbf91.pdf](https://www.isi2023.org/abstracts/ottawa-2023_488c130caaadd379dbf45084b6ecbf91.pdf)

Files/Uploads

[ML for subsidized housing- Chile](#)

Brief Description

Each trimester, the Central Bank of Chile calculates the Household Price Index.

This publication serves as an indicator of the price trends of the country's housing market.

The index is based on administrative records of housing transactions, which include various types of properties.

To ensure the accuracy of the index, transactions that are part of fully subsidized social housing complexes must be identified and excluded.

This paper addresses this need by proposing a supervised classification approach as a binary statistical classification problem using Machine Learning Models.

The chosen model, a Random Forest Model with measures to prevent overfitting, achieves a high rate of recall and accuracy in detecting social housing transactions.

Results show that from 2004 to 2022, 9% of all transactions in Chile corresponded to social properties.

Furthermore, our study highlights the potential of machine learning for automating data processing in Central Banks, which can lead to enhanced accuracy in the creation of official statistics.

Abstract

Each trimester, the Central Bank of Chile calculates the Household Price Index. This publication serves as an indicator of the price trends of the country's housing market. The index is based on administrative records of housing transactions, which include various types of properties. To ensure the accuracy of the index, transactions that are part of fully subsidized social housing complexes must be identified and excluded. This paper addresses this need by proposing a supervised classification approach as a binary statistical classification problem using Machine Learning Models. The chosen model, a Random Forest Model with measures to prevent overfitting, achieves a high rate of recall and accuracy in detecting social housing transactions. Results show that from 2004 to 2022, 9% of all transactions in Chile corresponded to social properties. Furthermore, our study highlights the potential of machine learning for automating data processing in Central Banks, which can lead to enhanced accuracy in the creation of official statistics.