# On the influence of model fragment properties on a machine learning-based approach for feature location

Manuel Ballarín[1], Ana C. Marcén[1], Vicente Pelechano[2], and Carlos Cetina[1]

1 SVIT Research Group, Universidad San Jorge, Zaragoza, Spain

{mballarin,acmarcen,ccetina}@usj.es

2 VRAIN Instituto Universitario Valenciano de Investigación en Inteligencia Artificial, Universitat Politècnica de València, Valencia, Spain

pele@dsic.upv.es

**Palabras Clave**: Model Fragment Location, Feature Location, Machine Learning, Learning to Rank.
**Lugar de publicación**: Information and Software Technology, Volume 129, p. 106430, 2021.
**Índice de impacto**: JCR 2,73 - Cuartil Q2 - Posición: 31/108 - Área: Computer Science / Software Engineering
**DOI**: https://doi.org/10.1016/j.infsof.2020.106430

**Resumen(Abstract).** *Context:* Leveraging machine learning techniques to address feature location on models has been gaining attention. Machine learning techniques empower software product companies to take advantage of the knowledge and the experience to improve the performance of the feature location process. Most of the machine learning-based works for feature location on models report the machine learning techniques and the tuning parameters in detail. However, these works focus on the size and the distribution of the data sets, neglecting the properties of their contents.

*Objective*: In this paper, we analyze the influence of three model fragment properties (density, multiplicity, and dispersion) on a machine learning-based approach for feature location.

*Method*: The analysis of these properties is based on an industrial case provided by CAF, a worldwide provider of railway solutions. The test cases were evaluated through a machine learning technique that uses different subsets of a knowledge base to learn how to locate unknown features.

*Results*: Results show that the density and dispersion properties have a direct impact on the results. In our case study, the model fragments with extra-small density values achieve results with up to 43% more precision, 41% more recall, 42% more F-measure, and 0.53 more Matthews Correlation Coefficient (MCC) than the model fragments with other density values. On the other hand, the model fragments with extra-small and small dispersion values achieve results with up to 53% more precision, 52% more recall, 52% more F-measure, and 0.57 more MCC than the model fragments with other dispersion values.

*Conclusions*: The analysis of the results shows that both density and dispersion properties significantly influence the results. These results can serve not only to improve the reports by means of the model fragment properties, but also to be able to compare machine learning-based feature location approaches fairly improving the feature location results.