Machine learning model analyzes why couples break up
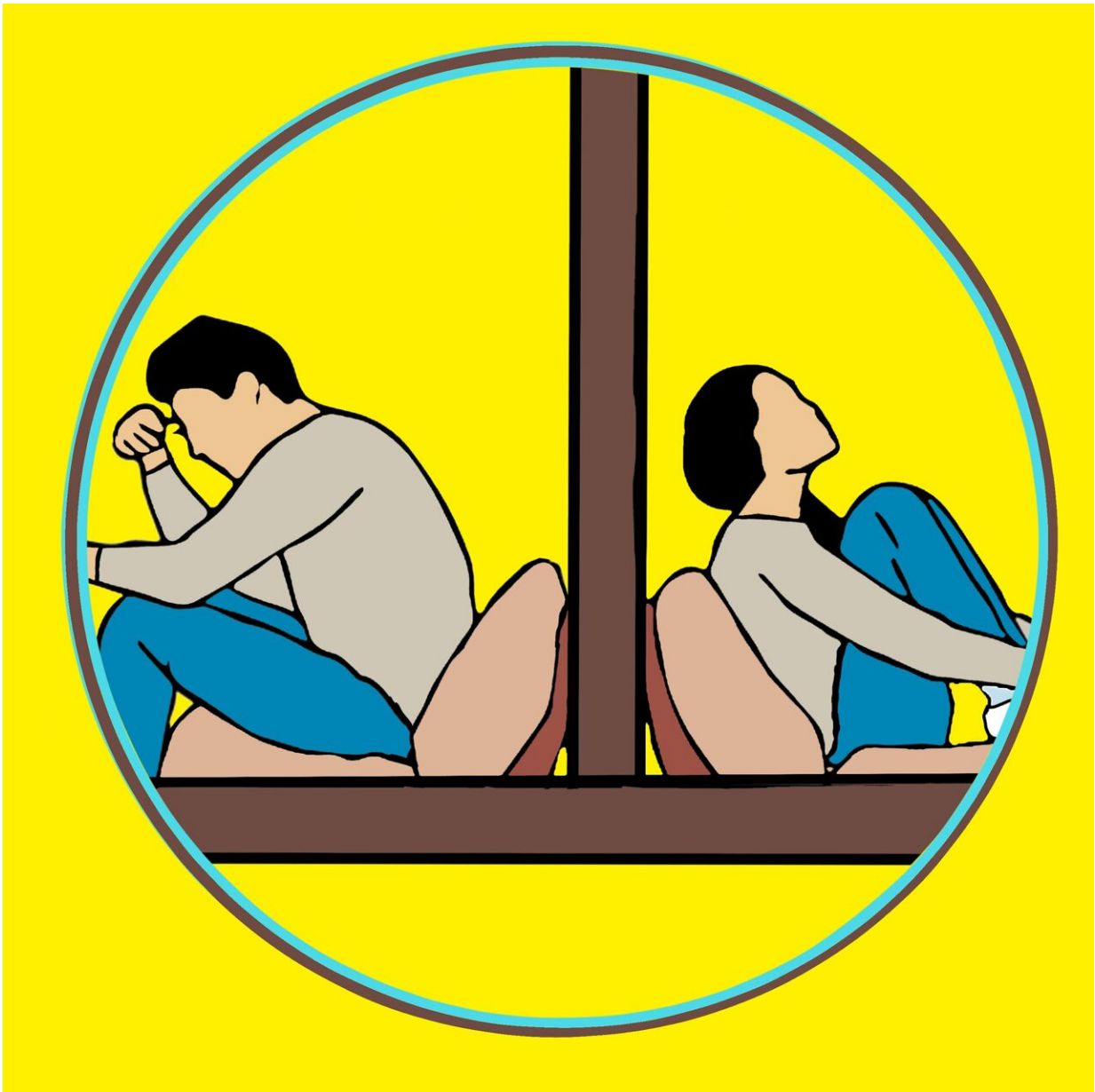
May 17 2023, by Letizia Mencarini

What does artificial intelligence offer that goes beyond traditional statistical models, such as regression analysis, to investigate the behavior of households, in particular the factors that cause the separation of couples and dissolution of the conjugal bond?

With Bruno Arpino (University of Florence) and Marco Le Moglie (Catholic University of Milan) we have analyzed data for over 2,000 German married or cohabiting couples, who were followed for a dozen years on average by the annual GSOEP survey (German Socio-Economic Panel), with more than 900 ending in separation.

By adopting a machine learning approach (specifically, Random Survival Forests) the procedure found on its own the relationship between the various factors contained in the database. In this case it considered more than 40 factors, from age to education level, from health to psychology traits: the mass of raw data was fed to ML, without making precise hypotheses, but simply indicating as event of interest the break-up of the union, and the algorithm indicated the influence of each factor contained in the data. The variables that pose the greatest threat to the stability of a union have been identified with an accuracy of 70% (a predictive ability that outperforms the 50% achieved by traditional regression methods).

Not only was ML able to discover the factors behind the breakup of couples, but it was also able to use this knowledge to predict the end of a union before it happens. This is also because, instead of submitting all the data available to an ad hoc algorithm, half were used to instruct the algorithm itself and the validity of the results was verified with the other half of the dataset.

The results of the analysis are very interesting, above all because the ML methodology is able to weigh the relative importance of various factors in causing the breakup. Factors that had been particularly influential in previous studies have instead lost their relevance here, like unemployment, and the partner's high level of education and income.

The four major risk factors, that emerged from the study are in descending order: personal satisfaction, the woman's quantity of paid work, some [personality factors](link) and age.

The strongest predictor of separation is personal satisfaction: if both partners are dissatisfied, obviously the couple won't last. Less obvious is that a strong drop in conjugal stability emerges when the woman is very satisfied with the union but the man much less so, while the reverse effect is less evident. If the woman works many hours outside the home, the risk of separation or divorce is higher, even when the man is more involved in domestic chores (but this result is nothing new and according to the existing literature it depends on the greater agency and independence of working women.

As for [personality traits](link), high extraversion in men (classically linked to a higher infidelity) and low openness in women, less adaptable to the changes brought about by cohabitation, are the traits that more strongly associated with the end of a couple. Also a low level of conscientiousness in both partners (understood as organizational capacity in daily life, and therefore—if low—as disorder and inability to respect commitments) does not help to stay together. But also a too high or too low level of neuroticism can be a problem. This result can be interpreted as the fact that suffering from excessive anxiety, jealousy, guilt, worry or anger clearly complicates the relationship.

This is true above all for women, but, on the other hand, those who don't feel this type of emotion could lead their partners to read that [personality](link)

trait as lack of interest (men, in this case). However, no pairing of personalities was determined that is more strongly associated with the breakup of the relationship. Finally, considering age, very young couples tend to be more unstable, but for women stability in relationships intensifies after the age of 40, while this is not the case for men.

ML analysis is not without limitations. In this case a major one is that it refers only to Germany and also has few details on the psychological aspects of the two partners. However, from a methodological point of view, the study demonstrates the great potential of ML techniques in demographic and sociological research in general, highlighting their ability to monitor and analyze a large number of predictive factors, to automatically find linear or non-linear relations, additive or non-additive relations between these factors and the outcome of interest, with greater precision and more robustness of estimates against collinearity than commonly used methods.

Provided by Bocconi University