ISSN 2281-4299

DIPARTIMENTO DI INGEGNERIA INFORMATICA
AUTOMATICA E GESTIONALE ANTONIO RUBERTI

SAPIENZA
UNIVERSITÀ DI ROMA

Marriage or Cohabitation? Using machine-learning techniques to predict the relationship choices in France, Germany and Italy.

Julia Nawaro
Cinzia Daraio

Technical Report n. 06, 2021

Marriage or Cohabitation? Using machine-learning techniques to predict the relationship choices in France, Germany and Italy

Julia Nawaro and Cinzia Daraio

Department of Computer, Control and Management Engineering A. Ruberti (DIAG)

Sapienza University of Rome

Abstract

This paper analyses the decline in marriage rate and the growing popularity of relationships alternative to marriage, which are a demographic and social change taking place in developed countries since the mid-20th century. We use machine-learning techniques including Logistic Regression, Support Vector Classifier, Random Forest and Gradient Boosting to analyse a rich dataset on Generations and Gender Programme for France, Germany and Italy. Data are at the micro-level and come from two waves of questionnaire (2003-2005 and 2007-2009). We develop a model to predict the relationship status (married vs cohabiting) over three years, taking into account the variables that influence the choice of marriage over cohabitation. Age, number of children, country of birth, employment status and views on marriage related topics are significant factors that influence the relationship status. Our model is able to predict the relationship status with over 85% accuracy for France, 82% for Italy and 80% for Germany.

Keywords: Marriage, Cohabitation, Family Formation, Second Demographic Transition, machine-learning techniques

1. Introduction

At the end of the 19th century, French researchers noticed significant demographic changes taking place in their country since the previous century, which was mainly reflected in a decrease in the average number of children per woman in reproductive age (Van de Kaa, 2002). The decision to have fewer and fewer children was a kind of revolution. At the end of World War II, researchers from the United States became interested in this phenomenon, already taking place in most highly developed countries (Lesthaeghe, 2011) and called this decrease in the average number of children as a demographic "transition" (Van de Kaa, 2002). Notestein (1945) emphasized the imbalance of the long-term demographic equilibrium determined by the lower level of birth-rate combined with a decline in mortality. It was therefore necessary to create a new balance with low levels of mortality and fertility. Notestein (1945) stated that the demographic transition is a common phenomenon in developed countries: each country, after reaching a certain level of development, will have to go through it.

However, there are not studies, to the best of our knowledge, that predicted another such sudden, widespread and rapid decline in fertility levels in European countries around 1965, much below the value ensuring the replacement of generations. The highest value of the Total Fertility Rate in EU-27 countries in the period 1960-2020, which was equal to 2.62 children per woman of childbearing age, was reached in 1964. It fell below 2 for the first time thirteen years later (1977). In 1986, the year when the concept of the second demographic transition was formulated, the fertility rate was already 1.71, and in 2018 only 1.54 (World Bank, 2020).

From the end of the 19th century, the average age of getting married for the first time was gradually decreasing and it reached its lowest value in the 1960s (Lesthaeghe & Neels, 2002). From 1965, the trend began to reverse, which is considered as another manifestation of the second demographic transition. The increase in the median age of starting a marriage resulted, among others, from the growing popularity of cohabitation as a premarital phase and living apart together relationships, as well as the later age of leaving family home. Additionally, more and more people generally preferred to live without a partner (Lesthaeghe & Neels, 2002).

Other manifestations of the second demographic transition were the increase in extra-marital births and the development of the *secularization*, or disassociation from religious concerns. The latter also contributed to the increase in the divorce rate and the critical consideration of a lifelong relationship (Lesthaeghe & Neels, 2002).

The theory of marriage developed by Becker (1973) completes the picture of the dissemination of informal relationships, which are characterized, according to his view, by lower exit costs and greater overall flexibility in comparison to marriage. This theory is based on the assumption that marriage is voluntary. It is then possible to apply the theory of preferences, according to which each person deciding to get married expects an increase in his/her utility. In order to achieve their utility, people try to find the best partner and thus compete with each other, which is why Becker believed that there is a "matrimonial market". The benefits of marriage should be balanced with the costs of looking for a suitable partner (mainly spending free time) and the organization of the ceremony itself (financial dimension). Becker postulated that the net profit from marriage is the greater the more complementary the goods provided by each partner. For when they are substitutes, the household is able to produce exactly as much final product as the partners separately, so the profit from marriage is zero. It was emphasized the importance of having children as the main goal of marriage, hence people who want relatively few kids would get married later or would divorce sooner.

Nowadays, the cohabitation is quite a common practice, especially among young generations, it is socially acceptable, but it still has different functions depending on the country. In the south of Europe, for example in Italy, a small percentage of people cohabit and they are mainly living in more urbanized areas. In Central Europe, such as the western part of Germany or the Netherlands, cohabitation is considered as an attempted marriage, and usually couples decide to marry when a child arrives. After all, in states such as France, the eastern part of Germany, Austria and Norway, cohabitation is an alternative to marriage, and these countries have a high percentage of extra-marital births (Nazio & Blossfeld, 2003).

The diffusion of cohabitation in France happened in the earliest, at the end of the 1960s, mostly among students, then in the upper class, and almost a decade later it affected the blue-

collar workers (Köppen, 2010). For the half of couples cohabiting in 1970-1980, this kind of an informal relationship was only a trial of marriage, but from the beginning of the 1980s, it was increasingly popular as a separate type of relationship (Köppen, 2010).

The above change of the 1980s in the perception of cohabitation did not apply to the inhabitants of West Germany, where cohabitation is still considered as a prelude to marriage. Since the last decade of the 20th century, premarital cohabitation has become a natural stage in the lives of people under 30 years old (Köppen, 2010). Being married and having children are related phenomena in Germany, so when a child is born, cohabitation usually turns into marriage. Indeed, in West Germany in 2004, only slightly more than 4% of mothers aged 35-39 lived in cohabitation, three times less than in the east of the country (Dorbritz, 2008). In the former Federal Republic of Germany, the percentage of married childless women was more than twice as high as in East Germany, which suggests a stronger position of the institution of marriage, not only as an aid in exercising parental authority.

In Italy, considering women born between 1977 and 1986 who experienced both childbearing and marriage, in 14% of cases the birth of the child preceded marriage (Istituto Nazionale di Statistica, 2019). Recently, intense changes have been taking place in Italy in the sphere of marriage and fertility. A particularly drastic change has taken place in the percentage of married women aged 25-34, which has more than doubled down in 2018 compared to 1981 (Istituto Nazionale di Statistica, 2018). Moreover, the proportion of divorced women over the age of 35 has increased significantly for each age group, while there have been more married and less widowed seniors in contemporary Italy.

In France, Germany and Italy, we can observe manifestations of the second demographic transition occurring in Western European countries. These three countries are characterized by different historical developments and distinct social background. However, from the 1970s, similar changes, including the decrease of marriage rate and the increase of the divorce rate, have been observed in all the three countries. In addition, women have also decided to have children later and later, as a result of, *inter alia*, wider access to contraception, extension of education time and greater activity in the labour market. It was also then that homosexual people who wanted to be able to formalize their relationship began to manifest for their rights. This led

to the creation of new regulations concerning living together, operating on different principles in the three surveyed countries. In France, both civil partnership (*Pacte civil de solidarité* - PACS) and marriage remain valid and are gender-neutral from 1999; in Germany, at present, the only formal union is marriage that can be entered into by heterosexual and homosexual persons (homosexual marriage replaced homosexual civil union in 2017); in Italy, marriage is possible only to heterosexual couples, and civil partnership only to homosexual couples (since 2016). Civil partnership places some of the marriage rights and obligations on couples remaining in the union. In return for the declaration of cohabitation and emotional support, the partners mainly receive financial benefits, i.e. the possibility of joint tax settlement, tax reliefs as well as the right to information when the partner is in hospital. The big difference between civil partnership and marriage is the right to adoption. In Italy, same-sex couples cannot adopt a child, so civil union does not regulate this issue. Spouses in France and Germany have the right to joint adoption, and thus both exercise parental authority and are jointly responsible for the debts caused by the children's education. PACS partners retain their rights in this respect as people in an informal relationship - only individual adoption is possible, i.e. only one person exercises power in the legal sense. In France, the changes in social sphere began to take place at the earliest and relatively uniformly across the country. The great political and social differences in contemporary Germany, which existed as two separate states for 40 years in the second half of the 20th century, caused a certain regional dichotomy in the pace and dimension of social change. In Italy, also characterized by cultural differences between the North and the South, the process of the second demographic transition began to take place at the latest and intensified changes started only at the end of the 20th century.

In this paper, we investigate the choice of relationship status (married vs cohabiting) to identify the most significant factors affecting this choice, and provide an empirical analysis on the three largest countries of the European Union in terms of population that are France, Germany and Italy. The selection criterion was also a large differentiation of their political and historical situation, which could have an impact on social tendencies, as the study covers liberal France, capitalist Germany and traditional Italy. In the mid-1990s, considering 25-29 years old women,

about one in four lived in a cohabitation in France, one in six in Germany and one in twenty in Italy (Di Giulio & Rosina, 2007).

The paper is organized as follows. The next section presents the main aim and contribution of the work. Section 3 reports an outline on the related literature. Section 4 describes the data analysed. Section 5 outlines the machine learning methods applied. Section 6 reports and discusses the obtained results, while Section 7 concludes the paper.

2. **Main aim and contribution**

The main goal of this paper is to identify the determinants of the choice of the marriage over cohabitation by developing supervised machine leaning models to predict the relationship status in the next three years in France, Germany and Italy. In this study, we consider a *single person* as a person who has no partner or has a non-stable partner; a *cohabiting person* as a person living with his/her partner in an unregistered or registered cohabitation and a *married person* as a person who has married another person of the same or different sex. These definitions are coherent with the information provided in the database used for the empirical analysis. Knowing current social trends is helpful to establish legal regulations answering real population needs. The observation and prediction of trends in the sphere of family creation may allow for the adjustment of legal regulations, as well as pro-family policy to the real needs of societies.

3. **Related literature**

The changes in marriage and fertility over the last half-century have been a rich area of research. In this section, we outline the existing literature on the determinants of the choice of the relationship type by citizens of European countries. The results may be sometimes contradictory, and the discrepancy is largely due to the cultural differences of the countries and their regions. The examined factors can be organized in four sub-groups, which are (i) age, cohort and gender;

(ii) education, situation on the labour market and wealth; (iii) previous marital status and number of children; (iv) situation in the family home and place of residence.

The influence of age and cohort factors is quite intuitive and consistent in the literature. Toulemon (1996) studied the increasing popularity of cohabitation in France. He created a model to analyse the changes in the factors influencing the decision to get married for women born in 1944-1968 of a similar age (under 35). His sample size reached over 40 000 people who were interviewed in 1994. Toulemon (1996) linked the decrease in the number of marriages of each generation with the decrease in the number of first marriages concluded by persons under 30 years old. As the cause of the phenomenon, Toulemon (1996) indicated the increase in the prevalence of cohabitation among the later born cohorts, which related to their transformation in worldview and hierarchy of values accompanying the second demographic transition. He found that women living in cohabitation for up to two years had three times greater chance of formalizing a relationship than women who were not in a relationship. With each successive year of living together, the probability of getting married decreased, approaching its value referring to women living without a partner, because cohabitation was becoming an alternative to marriage.

The influence of pregnancy on the formalization of the relationship varied depending on the length of the relationship. In the case of a woman who was in the first non-cohabiting relationship, pregnancy increased the probability of getting married seventeen times in relation to a non-pregnant woman who also had no partner before. In the case of cohabiting couples, pregnancy increased this chance a little more than three times. After the delivery, the probability of getting married by a French woman decreased to the same level as for non-pregnant and childless women.

Nazio and Blossfeld (2003) examined the phenomenon of cohabitation among young women in Italy and Germany. They retrieved data from Fertility and Family Survey and obtained 3 234 observations for Italian, 2 555 for East German and 2 497 for West German women born between 1954 and 1973. The data were collected in the period 1992- 1996. The authors found that the cohabitation spread out quickly among younger cohorts in Germany. Although it was still an uncommon practice for Italian women, in all of three considered countries (France,

Germany and Italy), a monotonic decrease in the probability of getting married among younger cohorts was noted for women born in 1954 as the reference category. Nazio and Blossfeld (2003) observed that marriage and cohabitation were alternative forms of starting life together with a partner for younger women. They found also that completing a compulsory education had a significant effect on getting married and starting a cohabitation in Italy and West Germany. Furthermore, they found a significant positive effect of being employed on the choice of cohabitation in Italy.

The determinants of first relationship type choice in the Mediterranean countries - Spain and Italy - were investigated by Pereiro et.al. (2014). The data were retrieved from 2006 Survey on Fertility, Family and Values which collected data for Spanish women and 2003 Multiscope Survey on Family and Social Subjects that covered Italian females. The final database counted 9 302 observations for Spain and 22 181 for Italy. Comparing women born in 1951-1960 with those born in 1971-1980, the chance of choosing cohabitation has almost doubled in Italy and is 3.4 times higher in Spain. At the same time, in both countries, the odds of getting married were 60% lower for younger cohort. In Spain, employed women had higher odds of being in any relationship. As a first type of a union, they chose cohabitation over marriage much more often. As Pereiro et.al. (2014) highlighted, this result stayed in line with other studies of many European countries. The relationship was completely different in Italy, where the chances of entering both formal and informal union decreased as a woman became employed, by 37% for marriage and 17% for cohabitation. The authors tried to explain this phenomenon by referring to a traditional Italian model, where male's role within a relationship is more significant and where family formation is induced by parents' economic support. Pereiro et al. (2014) calculated that in Italy and Spain the likelihood of getting married without previous cohabitation almost doubled when a woman was pregnant which confirmed that in these Southern European countries marriage and childbearing are interconnected demographic phenomena. They found also that the circumstances in which a woman leaves her family home had a significant influence on first relationship type. The likelihood of cohabitation was two times higher for women that had moved out from parents before first union formation in Italy and Spain. Moreover, women whose

parents lived separately had higher chances of choosing cohabitation over marriage as a first union than women from an intact family.

Kalmijn (2011) dealt with the influence of men's earnings and employment on the choice of union type. His study was based on panel data from 13 European countries (Denmark, Finland, Germany, Austria, the Netherlands, Belgium, France, Great Britain, Ireland, Portugal, Spain, Italy and Greece) collected in 1994-2001. The author limited the sample to men who were not married at the beginning of the study - their number was 17 743. In the analysed period, 4 492 of them decided to start a relationship, of which 2 499 - cohabitation and 1 993 - marriage; in turn, 1 498 men ended the cohabitation and got married. The collected data emphasized the attractiveness of working men on the marital market, whose chance of creating any type of relationship was 58% higher than that of the unemployed. Employment also increased the probability of getting married, which was 48% higher than cohabitation. The type of contract with the employer was important as well - a man working temporarily had 23% less chance of entering a marriage rather than cohabitation than a person who was employed on a permanent basis.

Bradatan and Kulcsar (2008) drew attention to the effects of demographic changes in Hungary two decades after the overthrow of socialism, i.e. reduced fertility, relatively high mortality and high migration rate. The number of divorces and couples living in cohabitation increased, which suggested that the countries of Eastern Europe entered the phase of the second demographic transition. Their research was based on data from the Gender and Generations Survey from 2001, and their areas of interest were the factors of entering into relationships (both formal and informal) and the diversity among cohorts. The cross-sectional data analysed in the study were made on a sample of 16 363 people aged 18-74 and used over 600 variables. For the purposes of the article, the authors limited themselves to a sample of women and selected variables related mainly to family formation and fertility. Over time, Hungarian women began to cohabit at a young age and did not decide to get married right away. The likelihood of getting married for the first time was five times lower for the cohort of women born in 1975 than for those born 30 years earlier, and the chance of a couple cohabiting before the union was formalized was seven times higher for the respective groups. Bradatan and Kulcsar (2008) concluded that the later a person decides to enter a cohabitation for the first time, the more

likely they will stay in it. This is mainly explained by the fact that couples who start cohabitation later are often divorced or widowed and do not want to marry again, either because of dissatisfaction with their previous marriage, or - in the case of widowers - from a sense of obligation to be faithful to their deceased partner. The research of Bradatan and Kulcsar (2008) also showed that 72.3% of Hungarian women born in 1935 who decided to cohabit for the first time were previously married. This percentage was only 7.2% for women born 30 years later. The preferences of the cohorts have changed, and cohabitation has become the first type of relationship. Hungarian women's pregnancy increased the probability of getting married instead of cohabiting. Despite the increase in the percentage of children born out of wedlock from 13% in 1990 to 34% fourteen years later, marriage and fertility were still closely related events, and cohabitation was more often chosen by women who postponed the decision to have a child. Bradatan and Kulcsar (2008) studied the influence of parents' divorce on the tendency of adult children to start a given type of relationship. Children from broken families more often decided to cohabit instead of getting married, for which there could be at least several explanations. For example, observing the breakdown of their parents' marriage in childhood, such persons could be more cautious, and thus prefer a flexible form of relationship. Furthermore, it was found that legal formalisation of a new kind of relationship can induce its familiarisation.

According to the research of Perelli-Harris and Gassen (2012), covering nine Western European countries, registered partnerships, present in the Netherlands and France, are treated most similarly to marriages and are mentioned in the largest number of nineteen chosen policy areas (in eighteen for the Netherlands and in sixteen for France), applying during the relationship, after its dissolution or death and in case of having children. These countries are followed by Norway and Sweden in terms of marriage and cohabitation harmonization. Austria's, England's and Spain's policies regarding relationship statuses provide obligations or rights only in some spheres. Germany has one law on cohabitation and Switzerland has no explicit law regarding cohabitation. Perelli-Harris and Gassen (2012) concluded that many among analysed countries attempted to give partners living together more legal rights and responsibilities but also pointed that it is questionable if law should regulate this kind of relationship given its still unknown role

(marriage trial or less binding partnership). Moreover, they stated that legal regulation of cohabitation led to its greater acceptance in the society and thus it is a more frequent practice.

In Table 1 we report an overview on the variables considered in the literature, affecting the decision to enter a particular type of relationship.

Table 1 *Summary table on the most relevant variables considered in the literature*

Variable	**Reference of the study**
Age, cohort and gender	Toulemon (1996), Nazio and Blossfeld (2003), Bradatan and Kulcsar (2008), Pereiro, et.al. (2014)
Education, situation on the labour market and wealth	Nazio and Blossfeld (2003), Pereiro, et.al. (2014), Kalmijn (2011)
Previous marital status and number of children	Bradatan and Kulcsar (2008), Toulemon (1996), Pereiro, et.al. (2014)
Situation in family home and place of residence	Bradatan and Kulcsar (2008), Pereiro, et.al. (2014)

4. **Data**

The Generations and Gender Programme is an open-access data source (available at: https://www.ggp-i.org/) launched by the United Nations Economic Commission for Europe in 2000. It provides longitudinal data about life and family dynamics of individuals in 20 developed countries. The Generations and Gender Survey (GGS) covers information from most basic, i.e. age, sex, marital status or number of children, through more specific, i.e. type of occupied dwelling, satisfaction with job or marital status of parents, to opinions on various family/work issues. The source of the data for our study were the first and second waves of the GGS questionnaire (versions 1.3 and 4.3) for France, Germany and Italy. The total number of 29 666 people aged 18-80 for wave 1 and 16 235 people for wave 2 (who have participated in the first wave) were surveyed in total, in the three analysed countries. The number of available variables, concerning, among others, health, marital status, fertility, education and professional activity,

was approximately 2,500 for each wave. The surveys were conducted from 2003 to 2005 and then from 2007 to 2009. Data were collected using the Paper-and-Pencil Interview, Computer Assisted Personal Interview and Computer Assisted Telephone Interview methods and the survey depositors were: National Institute of Statistics and Economic Studies, Federal Institute for Population Research and Italian National Institute of Statistics.

For the purposes of this study, a new categorical variable describing the relationship status was created, having the following three values: married; cohabitating (however defined, i.e. PACS, registered cohabitation or a union not registered anywhere); single or having a non-resident partner.

In the analysis, we select over a dozen of factors that may affect the choice of the relationship type on the base of the literature review summarized in Table 1 and our own reasoning. In order to obtain the most consistent, effective and complete data for all three countries, some variables had to be omitted, mostly due to missing values or presence of outliers. At the end, the following variables were considered:

- age,
- sex,
- household type (living alone, single parent, living with parents, couple with no children, couple with children, other),
- household size,
- activity status (unemployed, student or in training, employed or in parental/care leave, retired, ill/disabled/other),
- number of kids,
- total number of co-resident partners,
- being born in the country of interview,
- education level (tertiary/not tertiary),
- opinion: It is all right for a couple to divorce even if they have children.

Two additional variables were included for France and Germany, as there were not available for Italy:

- having a child from any of the previous partnerships,

- opinion: Homosexual couples should have same rights as heterosexual.

After limiting the data sets to people for whom all features were assigned, they consisted of:

- 9 829 (wave 1) and 6 312 (wave 2) observations for France,
- 9 407 (wave 1) and 3 049 (wave 2) observations for Germany,
- 6 191 (both wave 1 and wave 2) observations for Italy.

For all the analysed countries and waves, more women took part in the study - on average over 55%. In the period from first to second wave, the percentage of singles decreased and spouses - increased, reaching the highest growth of 9.5 percentage points in Germany. Moreover, the percentage of cohabiting increased in France and Italy but decreased in Germany, which may suggest that in the first two countries, cohabitation is increasingly becoming an alternative to marriage, not a transitional phase as in Germany.

5. **Methods**

We investigate the choice of relationship by applying a binary Logistic Regression and predictive multiclass models. By means of these techniques, we look for variables influencing the choice of marriage over cohabitation and try to predict what relationship status a given person will have in the next three years.

Logistic Regression in its simplest version is a statistical technique to model the binary response variable (see e.g. Kleinbaum et al., 2002). The explanatory variables can be binary or continuous. By this approach, we calculate the probability of an input belonging to output classes and classifies it to the class with higher value of this probability. The aim of the Logistic Regression model developed here is to examine the factors influencing the choice of marriage over cohabitation; therefore, singles were excluded from the analysis. We chose a logit model due to the binary nature of the dependent variable, representing marriage vs cohabitation. We assess the impact of eleven explanatory variables on the choice between marriage and cohabitation for France and Germany, and nine for Italy. We use the logistic model as a helpful baseline to identify significant variables for further analysis. McFadden's pseudo R^2 was 0.32, 0.33 and 0.23 for the models for France, Germany and Italy, respectively. At the significance level of 0.1%, four

variables turned out to be significant for France, five for Germany, three for Italy; at a significance level of 1% - one for Germany; at a significance level of 5% - four for France and two for Germany, while at a significance level of 10% - one for Germany and two for Italy. Factors having *positive impact* on the odds of getting married were:

- for France: number of children, being a woman, age;
- for Germany: being born abroad, household size, being a woman, age;
- for Italy: household size, age.

Variables that *negatively* affected these odds turned out to be:

- for France and Germany: number of partners, having a child from a previous relationship, being employed (in reference to being unemployed), agreement that it is all right for a couple to divorce even if they have children (in reference to not having opinion), agreement that homosexual couples should have same rights as heterosexual (in reference to not having opinion)
- for Italy: number of partners, agreement that it is all right for a couple to divorce even if they have children (in reference to not having opinion), being a student (in reference to being unemployed).

In the next part of the analysis, several prediction models are developed and compared. After that, our aim is to predict a person's relationship status (married/cohabiting/single) in the next three years in the most accurate way. Since the response variable has three values, the problem comes down to a multiclass classification. Multiclass classification can be solved by one of the three most common strategies i) transformation to binary, which involves performing multiple binary classifications; ii) extension from binary, in which binary classification algorithms adapt to predict more than two classes; and iii) hierarchical classification based on a tree architecture (Aly, 2005). In this analysis, mainly the second method will be used, the most popular algorithms of which are: Decision Trees, k-Nearest Neighbours, Naive Bayes, Neural Networks and Support Vector Machines (Aly, 2005). The distribution of classes is unbalanced (see Figure 1), so the evaluation metric used is Balanced Accuracy. By calculating the average of Recalls for each class (with recall = $\frac{TP}{TP+FN}$, where *TP* stands for *true positive* and *FN* stands for *false negative*), all of them have the same weight (while for Accuracy, the result depends mostly from the

performance of bigger classes), which makes this measure appropriate for imbalanced multiclass classification (Grandini et al., 2020). To achieve the goal of the analysis, the independent variables come from the first wave of the GGS questionnaire and are the same as for logistic regression. Moreover, the relationship status from the first wave becomes input variable to prediction models, while the target variable is the relationship status from the second wave of the survey.

Figure 1 *Relationship status of people in waves 1 and 2*

Own elaborations on GGS data.

After combining the questionnaire waves, the final data frames used for the prediction models included 6218 observations for France, 2950 for Germany and 5968 for Italy. The models were estimated separately for each country. The data were divided into train and test sets. The performance of the algorithms was tested on 33% of the data randomly extracted. Moreover, we used the 5-fold cross-validation technique for hyper-parameters tuning. We implemented grid search or randomized search depending on the number of hyper-parameters. The following

methods were implemented: k-NN, Decision Tree, Gaussian Naive Bayes, Support Vector Classifier, Random Forest and Gradient Boosting.

The method of k-NN clusters training data and classifies the observations from test data based on whether most of the k closest neighbours belong to a given group. Values from 1 to 100 for the hyperparameter were tested, and the best mean Balanced Accuracy was achieved by k = 7, k = 22, k = 6 for France, Germany and Italy, respectively. A greater number of hyper-parameters instead characterizes tree-based algorithms. In Decision Tree, the optimization was performed on maximum depth of the tree, number of features to consider when looking for the best split, minimum number of samples in a leaf node and minimum number of samples required to split an internal node in order to overcome the problem of overfitting, quite common for this algorithm. The Gini Index (GI = $1 - \sum_{i=1}^{n}(p)^2$, where p is the probability of belonging to a specific class) was used to measure the quality of a split.

The next algorithms implemented are Gaussian Naïve Bayes and Support Vector Classifier. Gaussian Naïve Bayes is based on the assumption of features pairwise conditional independence, which intuitively might be violated in the case of GGS data. There were no hyper-parameters optimized in this model. In Support Vector Classifier, a kernel trick was used whereby the algorithm transforms the data to be linear from a non-linear relationship. The tuned hyper-parameters were then the kernel coefficient γ and the regularization parameter C. The algorithm does not support the multiclass classification by default and "one-versus-one" approach was used to handle this problem, which consists of creating hyperplanes pairwise separating classes (in this case, three Support Vector Machines were constructed).

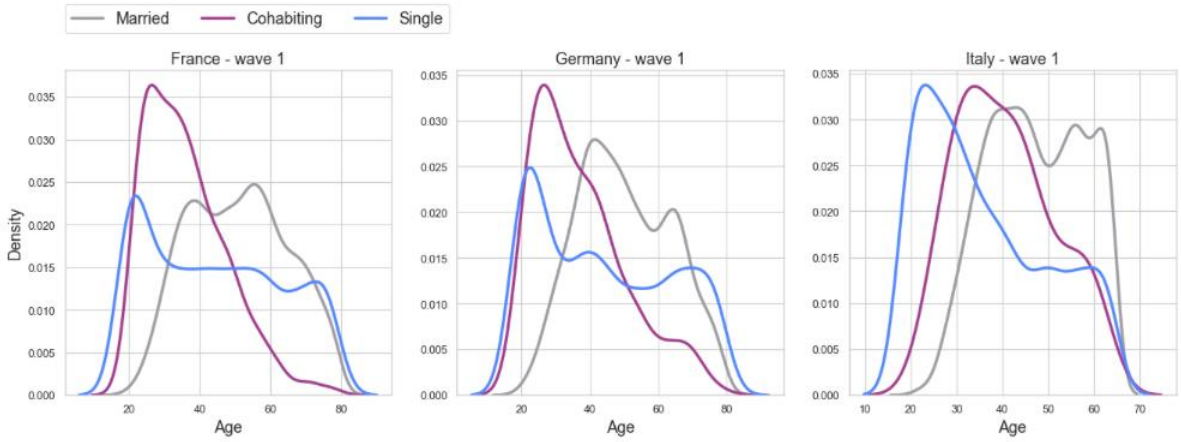
Finally, two algorithms of ensemble learning – Random Forest and Gradient Boosting - were implemented in order to try to outperform the classic algorithms by improving the accuracy and reducing the bias. Random Forest uses *bagging*, which consists of independent building of the learners and averaging their predictions, while Gradient Boosting uses *boosting*, in which the estimators are built consecutively. These techniques combine many learners (in this case - multiple Decision Trees) and form one stronger predictor. For cross-validation in Random Forest, the same hyper-parameters as in Decision Tree were tuned plus one additional, controlling the number of trees in the forest. Too long running time of Gradient Boosting prevented testing all

of the above hyper-parameters and only three of them were optimized. The optimal choices of the hyper-parameters are summarised in Table 4 in Appendix.

6. **Results and Discussion**

The obtained results show that France and Germany have similar age distributions for each type of relationship while Italy shows a different pattern (see Figure 2, first and second panels). In France and Germany, cohabiting is concentrated among young people, while married and singles are more dispersed. On the contrary, in Italy the age distribution of each relationship status is more concentrated in a given age group: singles are more concentrated in the youngest group, the older are more concentrated in those that choose cohabiting and those that choose marriage are concentrated in the oldest age group (see Figure 2, third panel). Consequently, the youngest Italians have a much higher probability of being single than living with their partner or being married than their peers from France or Germany, and *mutatis mutandis*, the oldest Italians have higher probability of being married than cohabiting or having no partner compared to their corresponding French and German groups.

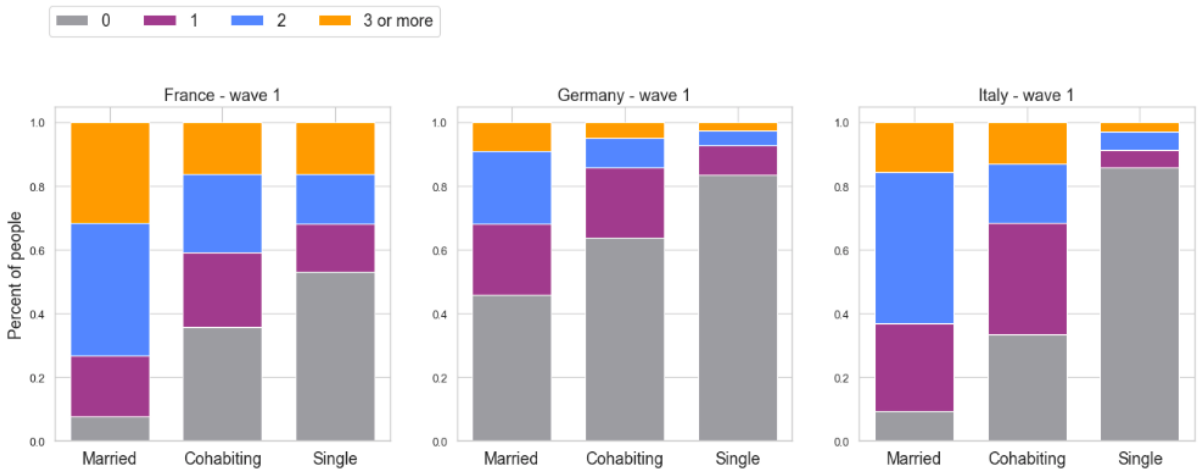
Figure 2 *Structure of married, cohabiting and single people by age (wave 1)*



Own elaborations on GGS data.

When considering the household types, there is a clear distinction between individuals in relationships and singles. The share of couples with children is the highest among Italian couples, followed by the German ones, while in France there are more couples with children among cohabiting couples than in married ones. This suggests once again that in France, cohabitation is an alternative to marriage. GGS data also show that in Italy, life of singles is different from that of their northern neighbours. The vast majority (59%) of Italian singles live with their parents, while in France and Germany singles are mostly living alone (69%). This is supported by the strong family ties among Italians, favouring marriage in the later choice of the type of relationship by young Italians. Spouses in general have children more often than singles or individuals in cohabitation. Among married couples with children in France and Italy, the model of two children clearly dominates against that of zero, one or three or more children, while in Germany the couples that have two children are in the same percentage that those with one kid (see Figure 3). Data from GGS confirm the strong phenomenon of childlessness among Germans, discussed by Dorbritz (2008), in the case of which over 40% of married people have no children, over four times more than in France or Italy. It is also worth noting the situation in Italy, where as many as two-thirds of people who live with a partner and are not married have a child. This suggests a profound change in the Italian traditional family model.

Figure 3 *Structure of married, cohabiting and single people in terms of number of children (wave 1)*



Own elaborations on GGS data

We considered also the changes in worldview that have been taking place since the 1970s summarized in Table 2. Regardless of the relationship status and country, the vast majority, on average 80% of people, believe that it is right for a couple to divorce even if they have children. This proves the rather developed idea of individualism and the lack of stigmatization of divorced persons by Western societies, which is an expression of the transformations of the 1970s. A greater discrepancy between the answers of spouses, cohabiting persons and singles occurs when asked about the rights of homosexual couples. With the statement that their rights should not differ from those of heterosexual people agreed only 33% of French spouses and 42% of German spouses. The percentage was higher by more than 10 percentage points for singles and by more than 20 percentage points for cohabites (see Table 2).

Table 2 *Opinions of married, cohabiting and singles on divorce and homosexual rights (wave 1)*

Opinion		France		
		Married	**Cohabiting**	**Single**
It is all right for a couple to divorce even if they have children	**agree**	79%	88%	81%
	no opinion	8%	6%	8%
	disagree	13%	6%	11%
Homosexual couples should have same rights as heterosexual	**agree**	33%	54%	46%
	no opinion	17%	17%	17%
	disagree	50%	29%	37%
Opinion		**Germany**		
		Married	**Cohabiting**	**Single**
It is all right for a couple to divorce even if they have children	**agree**	82%	91%	84%
	no opinion	12%	4%	9%
	disagree	6%	5%	7%
Homosexual couples should have same rights as heterosexual	**agree**	42%	65%	55%
	no opinion	19%	15%	17%
	disagree	39%	20%	28%
Opinion		**Italy**		
		Married	**Cohabiting**	**Single**
It is all right for a couple to divorce even if they have children	**agree**	66%	80%	69%
	no opinion	20%	10%	20%
	disagree	14%	10%	11%

Own elaborations on GGS data.

Table 3 summarizes the obtained results, which reports the intensity of variables that significantly increase or decrease the odd of being married.

The strongest positive influence on the odds of getting married was observed for the variable defining the *country of birth*. The odds of being married for people born outside Germany over the odds of being married for those born in Germany were equal 2.88 (so the odds are 188% higher for those born outside Germany than those born in the country, *ceteris paribus*). Given the variable with the second most effect (*household size*), the increase by a unit in household size causes 153% and 30% growth in the odds of being married for Italy and Germany, respectively.

The variable that has a negative impact on the odds of getting married, is *having one more partner* that causes the decrease of odds ratio by 60% for the French, by 64% for the Germans, and by 90% for the Italians.

Table 3 *Results: Intensity of variables that influence the odds of being married*

Variable/country	France	Germany	Italy
Not being born in the country of interview	---	2.876	---
Household size	---	1.297	2.534
Number of kids	1.356	---	---
Being female	1.187	1.290	---
Age	1.105	1.110	1.076
Agreement that homosexual couples should have same rights as heterosexual (reference category - no opinion)	0.7733	0.7581	---
Being a student or in training (reference category – being unemployed)	---	---	0.5907
Agreement that it is all right for a couple to divorce even if they have children (reference category - no opinion)	0.7054	0.5670	0.5855
Being employed or in parental/care leave (reference category – being unemployed)	0.6623	0.3830	---
Having children from previous partnership	0.2753	0.4833	---
Number of partners	0.4000	0.3578	0.1012

Own elaborations on GGS data.

An important aspect influencing the decision of getting married seems to be fertility. According to the regression results, having more children increases the odds of being married. It is assumed that the aim is to raise offspring in a full family, and the relatively high costs of a divorce may reduce the chances of parents breaking up. This factor can also be linked with age, because older people may have had time to father/give birth to more children, and moreover

the cohabitation is more popular among young people. In turn, the French and Germans get married less willingly when they already have offspring from a previous relationship. Thus, the data may indicate that those who already have a child are more cautious. In both countries, the study was conducted during a period of a relatively sharp increase in the percentage of extra-marital births. Unfortunately, it is not possible to conclude from the GGS data whether the previous relationship in which the child appeared was formal or not. The following thesis can be formulated - in the studied countries, decisions about marriage and fertility are more closely related phenomena in case of biological children than offspring in general. In addition, household size has a positive impact on the chances of getting married. While in Germany it may simply be associated with the number of kids and exhibit the relationship described above, in Italy, apart from children, living with parents/grandparents, whose more traditional views may lead young people to marry, may be important.

Another significant variable reducing the chances of getting married is *being employed* in both France and Germany and *being a student* in Italy compared to remaining unemployed. Students are mainly young people who, firstly, can only look for a suitable partner, secondly, they may simply not be able to afford the wedding, or thirdly, they may prefer to focus on themselves rather than entering into a serious, formal relationship. Being woman increases odds of being married in all three countries. According to Pereiro et al. (2014) and Kalmijn (2011), for Europeans, being employed most often increases the chance of entering into any type of relationship. For women this was also a premise for choosing cohabitation as the first type of relationship, and for men for choosing marriage in general. The employment impact is therefore quite complex and would require a more detailed analysis.

The odds of getting married in Germany increase when the person was not born in the country of interview. The net migration balance in Germany has remained positive for many years (Statistisches Bundesamt, 2020). Having the citizenship of the country in which one lives gives a number of rights that facilitate functioning in the society. Foreigners can obtain citizenship after 8 years of legal stay in Germany or after 2 years of marriage with a German (while staying legally in the country for at least 3 years). This can have a significant impact on relationship preferences.

The influence of the rest of the variables is quite intuitive: people with more co-resident partners choose a more flexible relationship, people with more liberal views choose cohabitation, while older people have a better chance of being married because their cohort was brought up in the times before the revolution of the 1970s. The results of the regression model seem to confirm the prediction of the literature reviewed in Section 3, who drew attention to the increasing prevalence of cohabitation as an expression of socio-demographic changes in the second half of the 20th century.

Optimal hyper-parameters values found via cross-validation were used to run the final prediction models. The Balanced Accuracy scores calculated on the test sets are summarised in Table 5 in the Appendix. For all the countries, the three best algorithms turned out to be: Support Vector Classifier, Random Forest and Gradient Boosting. The highest Balanced Accuracy was achieved for France (85.63%), then Italy (81.74%) and Germany (79.80%). The marital behaviour of the French is therefore the most predictable, which may mean that cohabitation is there an alternative to marriage for people with specific characteristics. In Germany and Italy, on the other hand, the position of cohabitation is still not precisely defined, which makes its prediction more difficult. Additionally, Germany and Italy are characterized by considerable socio-demographic differences between regions, which is not the case for France. Nevertheless, the existing literature reported in Section 3 suggests that the changes taking place in Italy are similar to those in France, while German society is more conservative with regard to the family formation.

A major limitation of the analysis is the fact that the survey did not distinguish civil union as a separate type of relationship and our study did not include some variables, e.g. the place of residence due to the missing data. Existing studies emphasized the importance of woman's pregnancy in choosing the type of relationship. Unfortunately, among surveyed women, none of them was expecting a child, therefore the variable was omitted. The question of sexual orientation remains too sensitive and was not included in the GGS study.

7. **Concluding Remarks**

In this paper, we analysed the socio-demographic behaviour of the inhabitants of Western Europe considering France, Germany and Italy. We developed prediction models to assess the structure of the population in short-term. This could be helpful to adjust the laws and policies in the considered countries. The situation is very dynamic, which will cause a decline in the fit of the models, the more so as the new generation (the so-called *Generation Z*), with different views, is entering the reproductive age, moreover the COVID-19 pandemic could have changed family preferences. We identified and explained the main determinants of the choice of the relationship type in three biggest in terms of population European countries. Their analysis may serve as the basis for the creation of legal regulations corresponding to the real needs of society. Such amendments could relate to financial security (joint tax settlements, inheritance) regulate the rights and obligations of parents living in a given type of relationship as well be a suggestion to tailor pro-family policies in order to achieve the demographic goal.

Investigating what are the factors that can affect the choice of the marriage associated with children could be helpful to sustain a series of policy in favour of increasing the birth rate needed in the Next Generation EU plans to recover Europe in the next years. Machine learning techniques have been helpful to identify such factors.

References

Aly M. (2005). Survey on multiclass classification methods. *Neural Netw*, *19*, 1-9.

Becker G.S. (1973). A Theory of Marriage: Part I, *Journal of Political Economy*, *81*(4), 813-846. https://doi.org/10.1086/260084

Bradatan C., & Kulcsar L. (2008). Choosing between Marriage and Cohabitation: Women's First Union Patterns in Hungary, *Journal of Comparative Family Studies*, *39*(4), 491-507. https://doi.org/10.3138/jcfs.39.4.491

Di Giulio P., & Rosina A. (2007). Intergenerational family ties and the diffusion of cohabitation in Italy, *Demographic Research*, *16*(14), 441-468.

Dorbritz J. (2008). Germany: Family diversity with low actual and desired fertility, *Demographic Research*, *19*(17), 557-598.

Gauthier, A. H., Cabaço, S. L. F., & Emery, T. (2018). Generations and Gender Survey study profile. *Longitudinal and Life course studies*, *9*(4), 456-465. http://dx.doi.org/10.14301/llcs.v9i4.500

Grandini M., Bagli E., & Visani G. (2020). Metrics for multi-class classification: an overview, *arXiv preprint arXiv:2008.05756.*

Istituto Nazionale di Statistica (2018), *Popolazione residente per stato civile*. Retrieved from https://www.istat.it/it/archivio/220713

Istituto Nazionale di Statistica (2019), *Matrimoni e unioni civili*. Retrieved from https://www.istat.it/it/files/2019/11/Report_Matrimoni_Unioni_Civili_2018.pdf

Kalmijn M. (2011). The Influence of Men's Income and Employment on Marriage and Cohabitation: Testing Oppenheimer's Theory in Europe, *Population*, *27*(3), 269–293.

Kleinbaum, D. G., Klein, M., & Pryor, E. R. (2002). *Logistic regression: a self-learning text*¸ Springer Nature Switzerland AG.

Köppen, K. (2010). *Marriage and Cohabitation in western Germany and France*. Universitätsbibliothek Rostock.

Lesthaeghe R. (2011). The "second demographic transition": a conceptual map for the understanding of late modern demographic developments in fertility and family formation, *Historical Social Research, 36*(2), 179-218.

Lesthaeghe R., & Neels K. (2002). From the First to the Second Demographic Transition: An Interpretation of the Spatial Continuity of Demographic Innovation in France, Belgium and Switzerland, *European Journal of Population, 18*, 325–360. https://doi.org/10.1023/A:1021125800070

Nazio, T., & Blossfeld, H.-P. (2003). The Diffusion of Cohabitation among Young Women in West Germany, East Germany and Italy, *European Journal of Population, 19*(1), 47–82. https://doi.org/10.1023/A:1022192608963

Notestein, F.W. (1945). Population - the long view. 37-57 in: T.W. Schultz (ed): *Food for the World*, Chicago, Chicago University Press.

Pereiro, T. G., Pace, R., & Didonna, M. G. (2014). Entering first union: the choice between cohabitation and marriage among women in Italy and Spain, *Journal of Population Research*, *31*(1), 51–70. https://doi.org/10.1007/s12546-014-9123-7

Perelli-Harris, B., & Gassen, N. S. (2012). How Similar Are Cohabitation and Marriage? Legal Approaches to Cohabitation across Western Europe, *Population and Development Review, 38*(3), 435–467. https://doi.org/10.1111/j.1728-4457.2012.00511.x

Statistisches Bundesamt (2020). *Migration between Germany and foreign countries*. Retrieved from https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration/Tables/migration-total.html

Toulemon L. (1996). La cohabitation hors mariage s'installe dans la durée, *Population, 51*(3), 675-715. https://doi.org/10.2307/1534489

Van de Kaa D.J. (2002). *The Idea of a Second Demographic Transition in Industrialized Countries*, Sixth Welfare Policy Seminar of the National Institute of Population and Social Security, 1-34.

World Bank (2020). *Fertility rate, total (births per woman).* Retrieved from https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=EU)

Appendix

Table 4 Optimal choice of hyper-parameters

Algorithm	**Technique**	**Results**		
		France	**Germany**	**Italy**
k-NN	grid search	k=7	k=22	k=6
Decision Tree	randomized search, n_iters=100	min_samples_split = 10 min_samples_leaf = 1 max_features: log2 max_depth = 20	min_samples_split = 5 min_samples_leaf = 4 max_features: sqrt max_depth = 10	min_samples_split = 5 min_samples_leaf = 4 max_features: sqrt max_depth = 15
Support Vector Classifier	grid search	gamma = 0.001 C = 1	gamma = 0.001 C = 1	gamma = 0.01 C = 1
Random Forest	randomized search, n_iters=100	n_estimators = 1000 min_samples_split = 10 min_samples_leaf = 1 max_features: log2 max_depth = 15	n_estimators = 1000 min_samples_split = 5 min_samples_leaf = 4 max_features: sqrt max_depth = 9	n_estimators = 1000 min_samples_split = 10 min_samples_leaf = 2 max_features: sqrt max_depth = 8
Gradient Boosting	randomized search, n_iters=10	min_samples_split = 10 min_samples_leaf = 4 max_depth = 9	min_samples_split = 10 min_samples_leaf = 4 max_depth = 50	min_samples_split = 10 min_samples_leaf = 2 max_depth = 6

Table 5 Balanced Accuracy scores for final models

France	
Algorithm	**Balanced Accuracy [%]**
Support Vector Classifier	85.63
Random Forest	84.89
Gradient Boosting	84.89
k-NN	82.60
Deicison Tree	82.35
Naive Bayes	77.56

Germany	
Algorithm	**Balanced Accuracy [%]**
Support Vector Classifier	79.80
Random Forest	78.49
Gradient Boosting	78.49
k-NN	75.06
Deicison Tree	74.64
Naive Bayes	64.42

Italy	
Algorithm	**Balanced Accuracy [%]**
Support Vector Classifier	81.74
Random Forest	81.74
Gradient Boosting	81.74
Decision Tree	81.18
k-NN	81.10
Naive Bayes	80.49