Supplementary Material B

This second part of the Supplementary Material is composed by a descriptive analysis from the three scenarios from the Section 5, as well as other results that demonstrate the better performance of the RM over the RF.

Table 1: Descriptive statistics from the predictors variable of the *Bolsa-Familia* database.

Variables	**Median**	**Mean**	**SD**	**CV**	**Min.**	**Max.**
CLR	1.235	1.594	1.427	89.562	0.130	10.970
DLR	0.925	1.221	1.125	92.105	0.130	7.140
Kx1	0.105	0.150	0.241	160.878	-0.500	1.160
Kx2	0.980	1.118	0.634	56.722	0.130	4.390
Kx3	0.915	1.215	1.127	92.741	0.130	7.140
Kx4	1.235	1.593	1.427	89.572	0.130	10.970
Kx5	1.825	2.096	2.093	99.879	-3.100	11.140

Table 2: Description of each category created in the study's questionnaire. The score evaluated in them give the information about the attendance of determined feeling in the transcription.

Predictor	**Description**
CAT1	Love as a relationship between specific people is not necessarily romantic
CAT2	Love as something essential/fundamental to human beings and/or humanity
CAT3	Love as something indefinable, ineffable, ethereal
CAT4	Love as something linked to sexual components
CAT5	Love as a romantic relationship between two people
CAT6	Love for family members
CAT7	Love as something geared towards family members
CAT8	Love as something geared towards friends
CAT9	Love towards divine or supernatural entities
CAT10	Love towards irrational animals
CAT11	Love towards inanimate beings
CAT12	Self-directed love
CAT13	Love as a source of positive emotions, attitudes and behaviors
CAT14	Love as a source of negative emotions, attitudes and behaviors

Analyzing the age distribution from the histogram presented in Figure 1 it can seem that for that study the distribution of age groups is asymmetric, where most of people are under their twenties.

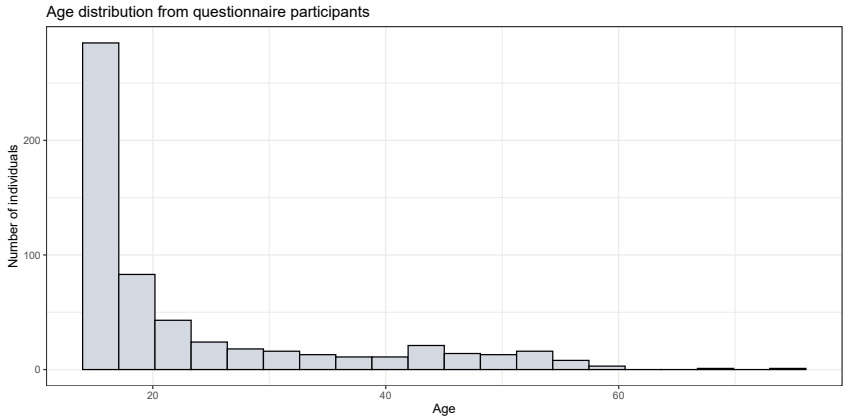

Figure 1: The histogram from ages show a degree of asymmetry between age groups.

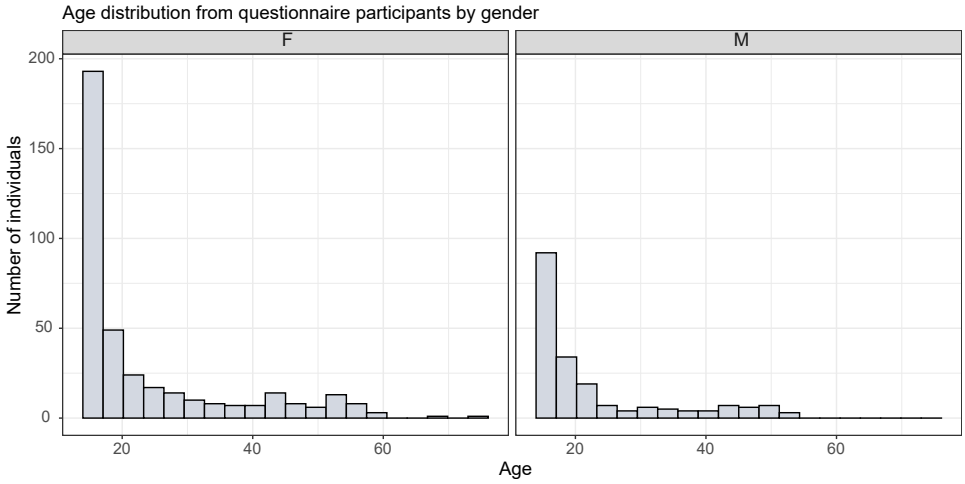

Figure 2: The histogram from ages by gender shows that the trend of younger people that attended the questionnaire.

Table 3: Summary statistics of independent variables from the model.

Covariate	Min	Max	Mean	SD
CAT1	0.000	4.000	0.138	0.409
CAT2	0.000	5.000	0.115	0.457
CAT3	0.000	5.000	0.198	0.557
CAT4	0.000	8.000	0.093	0.433
CAT5	0.000	4.000	0.299	0.578
CAT6	0.000	7.000	0.496	0.952
CAT7	0.000	2.000	0.232	0.435
CAT8	0.000	2.000	0.074	0.281
CAT9	0.000	2.000	0.041	0.208
CAT10	0.000	11.000	0.287	0.918
CAT11	0.000	4.000	0.041	0.253
CAT12	0.000	12.000	2.244	2.311
CAT13	0.000	8.000	0.365	0.933
CAT14	0.000	7.000	0.821	1.140
Age	14.000	73.000	23.830	11.808

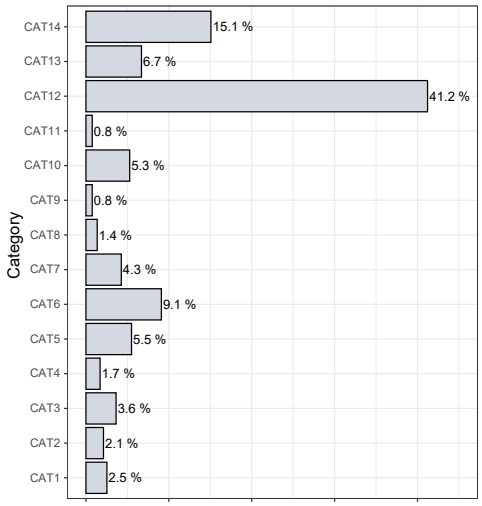

Figure 3: Percentage of responses in each category of love.

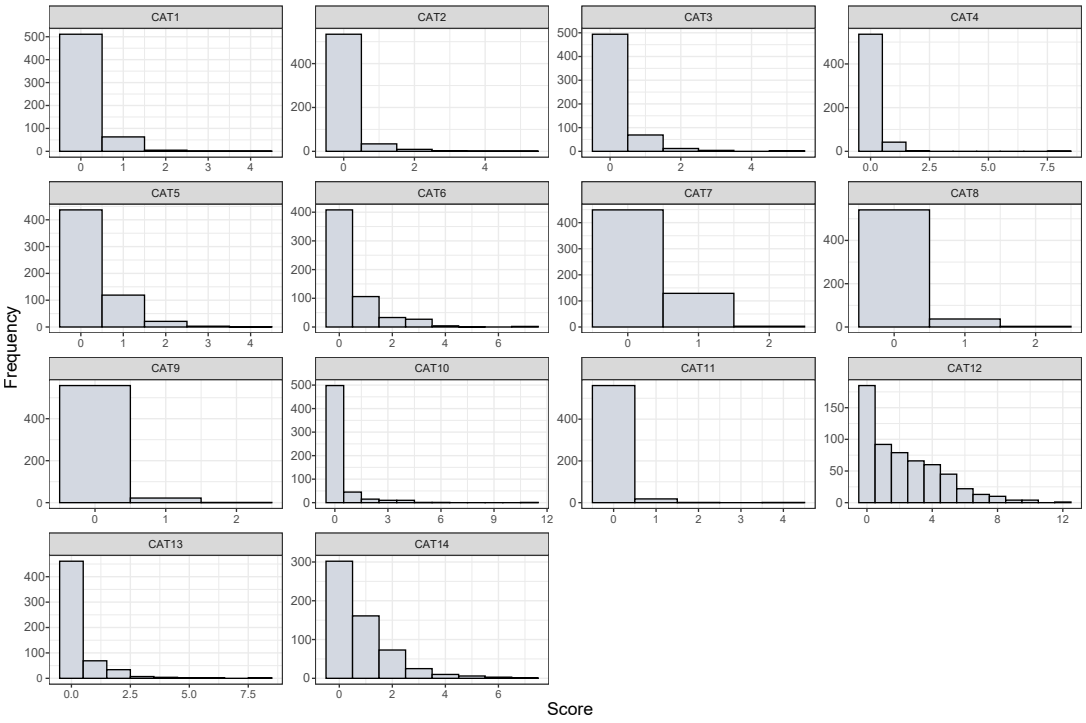

Figure 4: Histograms for each of the love categories.

Table 4: Variables that compose the third real-world application.

Variable	**Description**
y	Rate of use of social assistance program by municipality
COD_UF	Code to identify the Brazilian state that the city belongs
T_DENS	% of the population living in households with density greater than 2 people per bedroom
TRABSC	% of employed persons aged 18 or over who are employed without a formal contract
PPOB	Proportion of people vulnerable to poverty.
T_NESTUDA_NTRAB_MMEIO	% of people aged 15 to 24 who do not study or work and are vulnerable to poverty
T_FUND15A17	% of the population aged 15 to 17 with complete primary education
RAZDEP	Dependency ratio
T_ATRASO_0_BASICO	% of the population aged 6 to 17 years attending basic education that does not have an age-grade delay
T_AGUA	% of the population living in households with running water
REGIAO	Aggregation of states according to the regions defined by IBGE.

Table 5: Descriptive statistics from the continuous variable of the database.

Variable	**Median**	**Mean**	**SD**	**CV**	**Min.**	**Max**
y	0.095	0.101	0.057	0.561	0.001	0.246
T_DENS	23.065	25.127	13.000	0.517	0.650	88.640
TRABSC	24.755	25.226	9.853	0.391	3.030	62.230
PPOB	42.235	43.996	22.437	0.510	1.970	91.570
T_NESTUDA_NTRAB_MMEIO	13.670	14.777	8.751	0.592	0.000	55.250
T_FUND15A17	55.415	54.751	15.506	0.283	6.890	96.810
RAZDEP	49.595	51.488	8.949	0.174	29.170	118.040
T_ATRASO_0_BASICO	60.980	60.657	12.037	0.198	23.840	92.480
T_AGUA	90.280	85.596	14.722	0.172	0.150	100.000

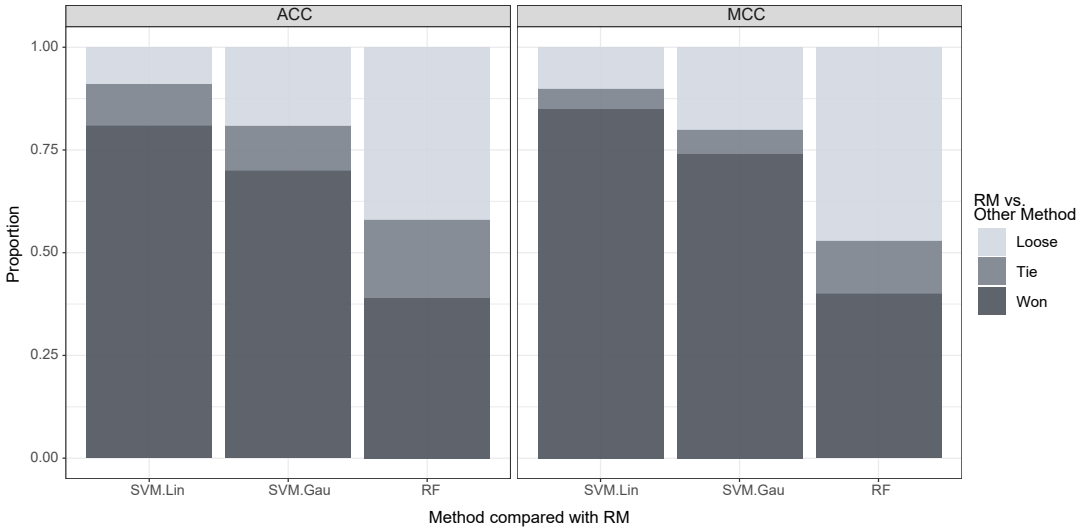

Figure 5: Proportion of the comparison status between random machines and the other methods. It can be noticed that for that task RM outperformed the single SVM algorithms and have an equivalent performance with respect to RF, in the first real-world application.

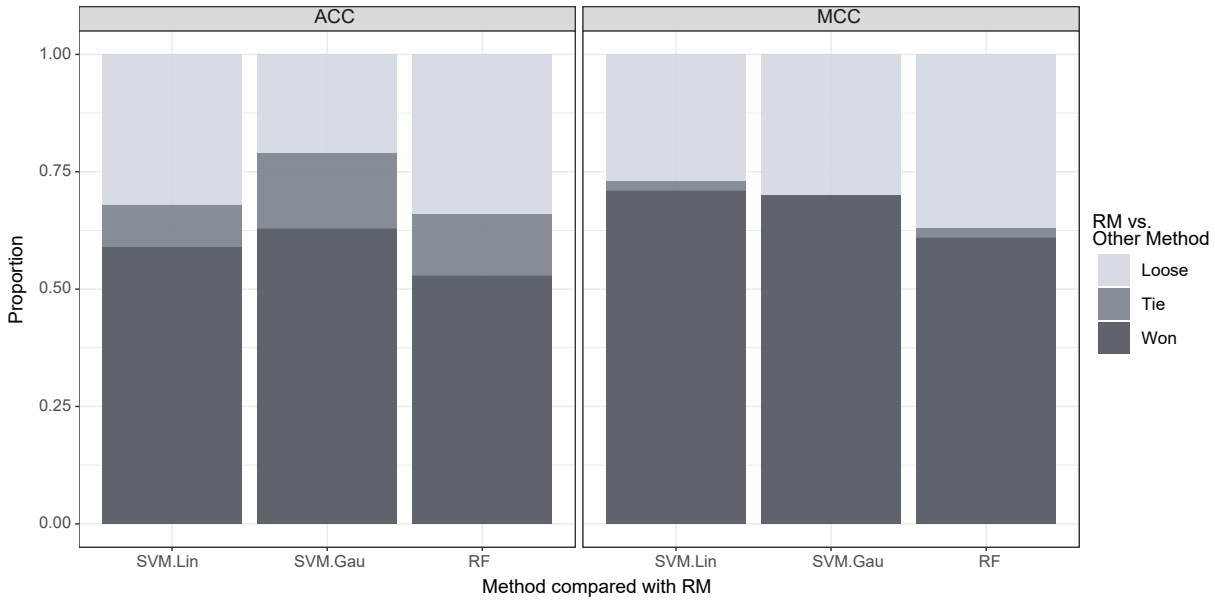

Figure 6: Proportion of win-tie-loose ratio between the random machines and the other models from the gender prediction.

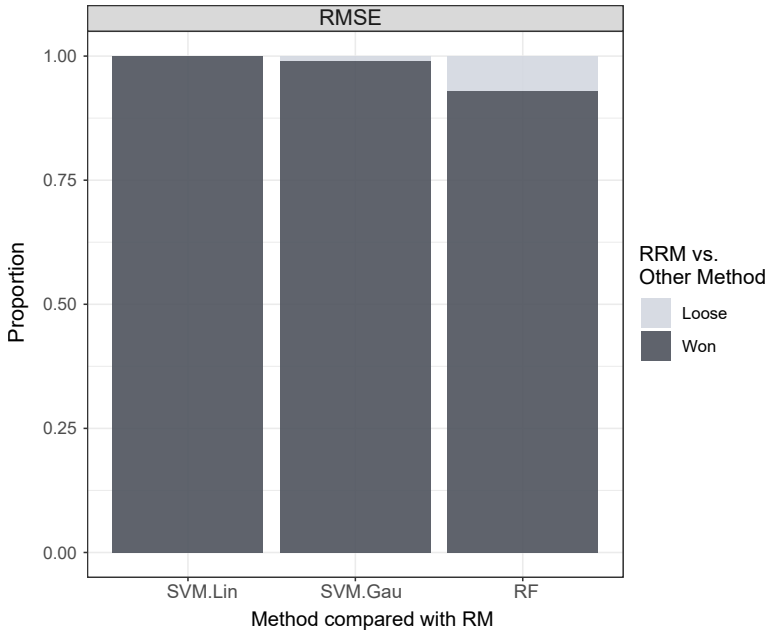

Figure 7: Proportion of the number of times that RM presented a lower RMSE than other methods, from the *bolsa familia* application.