**INTRODUCTION**

**Background** Internet voting in Switzerland started 2003 in one municipality of the Canton of Geneva only, expanded to several other cantons and was made available to expatriates as well. In a democracy such as the Swiss one, with several referendum votes held throughout the year, the introduction of internet voting not only leads to a debate on the digital divide but potentially affects the political system at large.

- Socialization and voting theories, the technology acceptance model as well as empirical experience from Estonia would suggest that internet voting became habitual over time. However, user rates of internet voters in Geneva did not increased so far.

**Goal of this study** Complimentary to prior research, a deep learning approach exploring to what degree neural networks can predict internet voting is suggested. With a variable importance assessment we want to find out why internet voting is not picking up in Geneva.

**DATA**

The Canton of Geneva maintains a centralized vote register identifying all citizens with a unique key. Data entries allow to track whether a person voted or not, and if yes, which of the three voting channels was being used, i.e. ballot box, postal, or internet voting.

**Variables in the data set**
- Categorical: voting channel (vtype), civil status (cstat), sex, municipality (munic), electoral district (eldis), professional situation (profsit) and status (prof) as proxies for education and income, i-voting registration requirement (regreq).
- Numerical: age, voting frequency (vcount), internet voting frequency (ivcount), internet voting opportunity (iopport).

**RESULTS**

**Prediction performance** While the random forest model achieved 76.6% accuracy, 81.6% is attained with a neural network (Fig. 1).

**Model interpretation** In order to obtain comparable results a model-agnostic interpretation of variable importance (VI) proposed by Fisher et al. (2018) was applied. VI is measured by calculating the decrease in prediction accuracy after permuting one feature at a time (Fig. 2). Due to entity embedding, categorical variables are more useful for the deep learning model than for random forest in general. However, besides composite voting behaviour indicators such as the cumulative frequency of voting or i-voting, age proved to be the most important individual socio-demographic characteristic for predicting the voting channel. Read: if age distribution would be random, the overall prediction performance of the neural network model would decrease by roughly 12%.

**CONCLUSIONS AND FUTURE WORK**

- In this so far most encompassing vote registry study on internet voting in Geneva the neural network model outperforms the random forest baseline model by 5%.
- Individual voting history and age are clearly more important for predicting future voting behaviour than features such as professional status or sex.
- For future work we would include additional variables such as the actual content of referendum votes as well as characteristics of the municipalities the voters live in.
- As a limitation we concede that time based factors were not explicitly taken into account for this analysis.

**LITERATURE**