

Figure 4: The example-based user interface for configuration-free data cleaning.

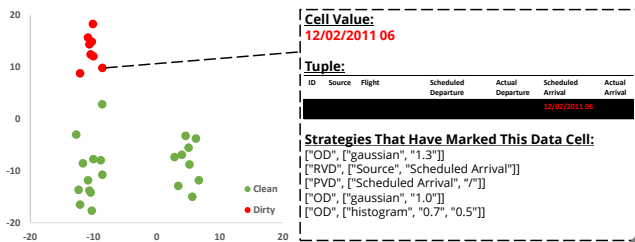


Figure 5: These 2D projected clusters contain either clean (green) or dirty (red) data cells for one column. By clicking on a point, the user can inspect the actual value and the error detection strategies that marked this particular data cell as a data error.

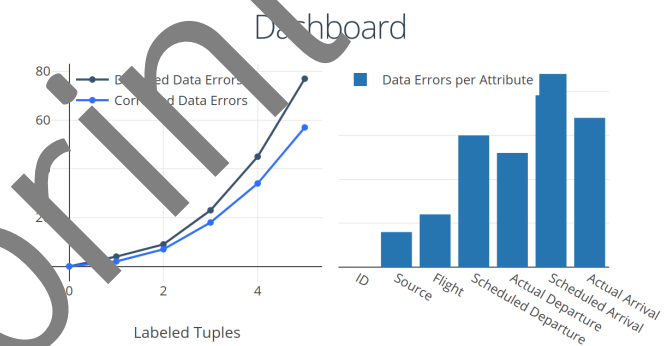
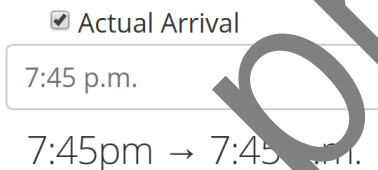


Figure 7: The user dashboard and data cleaning progress.



Model 1: **0.35** | Model 5: **0.17** | Model 12: **0.87**
 #Exact Matches: **3** | #Pattern Matches: **142**

Figure 6: The user can inspect the confidence of error corrector models for a particular correction and the number of exactly/approximately matched corrections from the Wikipedia page revision history.

configuration-free error detection system. In *SIGMOD*, pages 865–882, 2019.

- [15] S. J. Pan and Q. Yang. A survey on transfer learning. volume 22, pages 1345–1359, 2009.
- [16] C. Pit-Claudel, Z. Mariet, R. Harding, and S. Madden. Outlier detection in heterogeneous datasets using automatic tuple expansion. Technical Report MIT-CSAIL-TR-2016-002, CSAIL, MIT, 2016.
- [17] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré. Holoclean: Holistic data repairs with probabilistic inference. volume 10,

pages 1190–1201, 2017.

- [18] H. Schütze, C. D. Manning, and P. Raghavan. *Introduction to information retrieval*. Cambridge University Press, 2008.
- [19] L. Visengeriyeva and Z. Abedjan. Metadata-driven error detection. In *SSDBM*, pages 1–12, 2018.
- [20] P. Wang and Y. He. Uni-detect: A unified approach to automated error detection in tables. In *SIGMOD*, pages 811–828, 2019.
- [21] M. Yakout, L. Berti-Équille, and A. K. Elmagarmid. Don't be scared: use scalable automatic repairing with maximal likelihood and bounded changes. In *SIGMOD*, pages 553–564, 2013.
- [22] M. Yakout, A. K. Elmagarmid, J. Neville, and M. Ouzzani. Gdr: a system for guided data repair. In *SIGMOD*, pages 1223–1226, 2010.
- [23] Z. Yu and X. Chu. Piclean: A probabilistic and interactive data cleaning system. In *SIGMOD*, pages 2021–2024, 2019.