Inflated Count Data Model: Predictive Model for Indian Patent Renewal Life

A Thesis
Submitted in Partial Fulfillment of the Requirements for the Executive Doctoral Programme in Management

Indian Institute of Management Indore

by
Ashit Kumar
EDPM-2020
Operations Management and Quantitative Techniques

November 2024

Thesis Advisory Committee

Prof. Pritam Ranjan
[Chair]

Prof. Arnab Koley
[Member]

Prof. Mukul Gupta
[Member]

Abstract

In the world of innovation and technology, patents are an essential tool that promotes innovation (Bloom and Reenen, 2002; Leung and Sharma, 2021) and fosters the development of new ideas. Patents are an important aspect of the innovation ecosystem, providing legal protection and exclusive rights to inventors and innovators to prevent others from making, using, or selling their inventions for a specific period (patent life). Researchers have placed a significant emphasis on patent life estimation as it serves as an indicator of a patent's value, quality, and potential for future revenue generation. Patents have to be renewed every year by paying a renewal fee, and thus the life of a patent is calculated using the number of renewals. In the last few decades, many researchers have done studies on patent renewal data, but mostly they utilized patent renewal data from a patent valuation and patent quality perspective (e.g., Pakes, 1984; Pakes and Schankerman, 1984; Sullivan, 1994; Bessen, 2008; Svensson, 2012; Danish, Ranjan, and Sharma, 2020). Apart from research on patent value using such data, many researchers have utilized the survival model of patent renewal data to identify the determinants of patent value (Maurseth, 2005; Serrano, 2010; Danish, Ranjan, and Sharma, 2021a). An accurate prediction of patent renewal life not only helps in identifying the patent quality, and patent value indicator, but it also plays a significant role in patent acquisition decision. The predictive models of Indian patent renewal life have not been explored in the existing literature, and this thesis tries to address this research gap.

In India, the legal framework for patents is provided by the Indian Patents Act, 1970, which grants exclusive rights to the patentee for a period of 20 years from the date of filing. As per Indian Patent Office (IPO) data, in 2021-22, 66440 patents were filed, 66571 were examined and 30073 were granted. Note that the applications

examined in 2021-22 may have been filed in previous years[1]. Despite the significant number of patents granted, a large proportion of them expire at an early stage due to non-payment of renewal fees (Danish, Ranjan, and Sharma, 2020). We utilized patent level information data, which was collected from the IPO website[2] and PatSeer[3] for all granted patents filed between 1994 and 2005, for our study. The variable of interest for our research is the patent renewal life. The main objectives of this research work are : i) development of statistical and machine learning (ML) models to predict the renewal life of Indian patent and thus a indicator for patent value, patent quality ii) to identify the factors that influence the patent renewal life.

We investigated the state-of-the-art ML models such as decision tree, random forest, artificial neural network, support vector machine, XGBoost, and multiple linear regression. It turned out that the renewal life values from data lies in the range {0,1...20}, and exhibit inflated counts and 0 and 20. As a result, Further exploration on statistical model specifically proposed to tackle inflated count data was explored and predictive models were built using binomial regression, Zero and N Inflated Binomial (ZNIB) regression and a new proposed model-mixture model. None of these models turned out to be very effective in giving accurate predictions. We developed hybrid models for inflated count data that can combine the capability of two models — advanced machine learning classification algorithm, and generalized statistical regression algorithm. Predictive models were built using proposed hybrid model with three different classifier — random forest, support vector and XGBoost classification in conjunction with binomial regression model. The proposed hybrid model demonstrates significantly superior performance.

[1]https://ipindia.gov.in/writereaddata/Portal/Images/pdf/Final_Annual_Report_Eng_for_Net.pdf
[2]https://ipindiaservices.gov.in/publicsearch
[3]PatSeer - Gridlogics Technologies Pvt Ltd data.

The second objective of our research work was to identify the important variables impacting the patent renewal life and in turn the patent value, patent quality. This objective was achieved by extracting the relative importance of variables by utilizing permutation importance technique on machine learning models. Comparative permutation importance of variables from predictive models models clearly suggest that the top variables impacting patent renewal life are time gap between date of filing and granting date (grant lag), the number of jurisdiction the patent has been filed in (family size), patentee type, and whether the owner is domestics or foreign.

We developed two methodologies — mixture model and hybrid model. The mixture model is similar to the ZNIB model proposed by Sweeney, Haslett, and Parnell (2014) but it differs from their model as the ZNIB model assumes binomial source for all $y \in \{0, 1...20\}$ and additional source for $\{0, 20\}$ whereas the mixture model assumes unique sources for zero, N (= 20 in our case), and $y \notin \{0, 20\}$. Although mixture model performed at par with existing ZNIB model, there is scope for improvement in predictive model efficiency of the mixture model. We developed a hybrid model, by harnessing the benefits of machine learning classification algorithm and binomial regression algorithm. Hybrid model methodology concept can also be extended to other inflated common count data distribution for efficient predictive modeling.

A noteworthy aspect of this study is the predictability of patent renewal lifespans, which ultimately provides insights into patent value, patent quality and potential for future revenue generation. Apart from the development of predictive model the study also contributed to the literature of patent through the identification of the important patent characteristics of Indian patent renewal life. Our research work also contributed to the field of inflated count data modeling through the development of new methodologies.

Contents

6.3 Conclusion

The objective of our research work was to develop a methodology for efficient predictive models for inflated count distribution data. We developed a novel - Hybrid model approach by harnessing the capability of machine learning and statistical modeling for predictive modeling for count data inflated at zero and any integer N ($N > 0$). Our other objective of the study was to build a patent life predictive model for Indian patents. We built a patent life predictive model using binomial regression, support vector machine regression and hybrid model and computed model performance parameters. The predictive model's comparative result Table 6.9 and prediction error histogram Figure 6.8 and scatter plot Figure 6.9 suggests that hybrid models provided accurate predictions for inflated data points and outperformed standalone advanced machine learning and statistical modeling techniques in terms of predictive capability and computational time.

The proposed hybrid model approach for handling inflated data may be a novel approach as it combines the capabilities of machine learning techniques and statistical methods. Further, it can be extended to any other common count data probability distribution, such as the zero and N inflated Poisson distribution, zero and N inflated negative binomial distribution for efficient predictive modeling. Overall, this study provides a valuable contribution to the field of inflated count data modeling and

patent renewal life by demonstrating the effectiveness of statistical methods, machine learning technique, and hybrid models in predicting the renewal probability of Indian patents.

Chapter 7

Summary, Contribution and Research Implication

7.1 Thesis Summary and Conclusion

Researchers have placed a significant emphasis on Patent renewal life estimation as it serves as an indicator of a patent's value, quality, and potential for future revenue generation. An accurate estimation of patent renewal life not only helps in identifying the above indicators but it can also play a significant role in patent acquisition decisions. We investigated the following objective in this thesis

i) Development of an Indian patent renewal life predictive model to predict the

renewal life of the patent and an indicator for patent value, and patent quality.

ii) To identify the factors that influence Indian patent renewal life.

iii) To develop a new methodology for predictive modeling of inflated count data.

As discussed in earlier chapters these objectives were achieved by: i) Developing a patent renewal life predictive model using machine learning, statistical, and hybrid model techniques for Indian patent data. ii) Extracting feature importance of predictors from predictive models using permutation importance techniques; and iii) Developing an inflated statistical mixture model and a hybrid model by combining machine learning and statistical modeling capability for the modeling of complex inflated data. In this chapter, we will discuss the comparison of results of all the predictive model and research outcomes, conclusion, research contributions, and practical implications.

7.1.1 Patent Renewal life Modeling Results and Conclusion:

The first objective of our study was to build patent renewal life predictive models for Indian patents. The histogram of the variable of our research interest patent renewal life ('Renewalyear') of the Indian patent data set suggest its distribution close to binomial distribution with inflated data at zero and twenty (Figure 1.1). The collected Indian patent data was utilized to build optimized predictive models using

machine learning techniques such as decision tree, random forest, artificial neural network, support vector machine, XGBoost, and multiple linear regression, however, these advanced machine learning techniques did not perform efficiently at inflated data points zero and twenty. Further exploration on a statistical model specifically proposed to tackle inflated count data was explored and predictive models were built using binomial regression, zero and N inflated binomial regression and a new proposed model-mixture model. These statistical model results also revealed that these models could not perform at par with the ML algorithm on inflated count data. The unavailability of an efficient predictive modeling methodology for inflated count data suggested the development of a new methodology. We proposed a hybrid model for inflated count data that can combine the capability of two different models — an advanced machine learning classification algorithm and a generalized statistical regression algorithm that provides more efficient predictive results. Three predictive models were built using the proposed hybrid model with three different classifiers — random forest, support vector machine, and XGBoost in conjunction with a binomial regression model. RMSE and pearson's correlation values from all the models were extracted and a comparative table of these performance parameters is tabulated in Table 7.1. We also plotted the prediction error of test data from all the predictive models. A comparative prediction error plot was plotted from the best

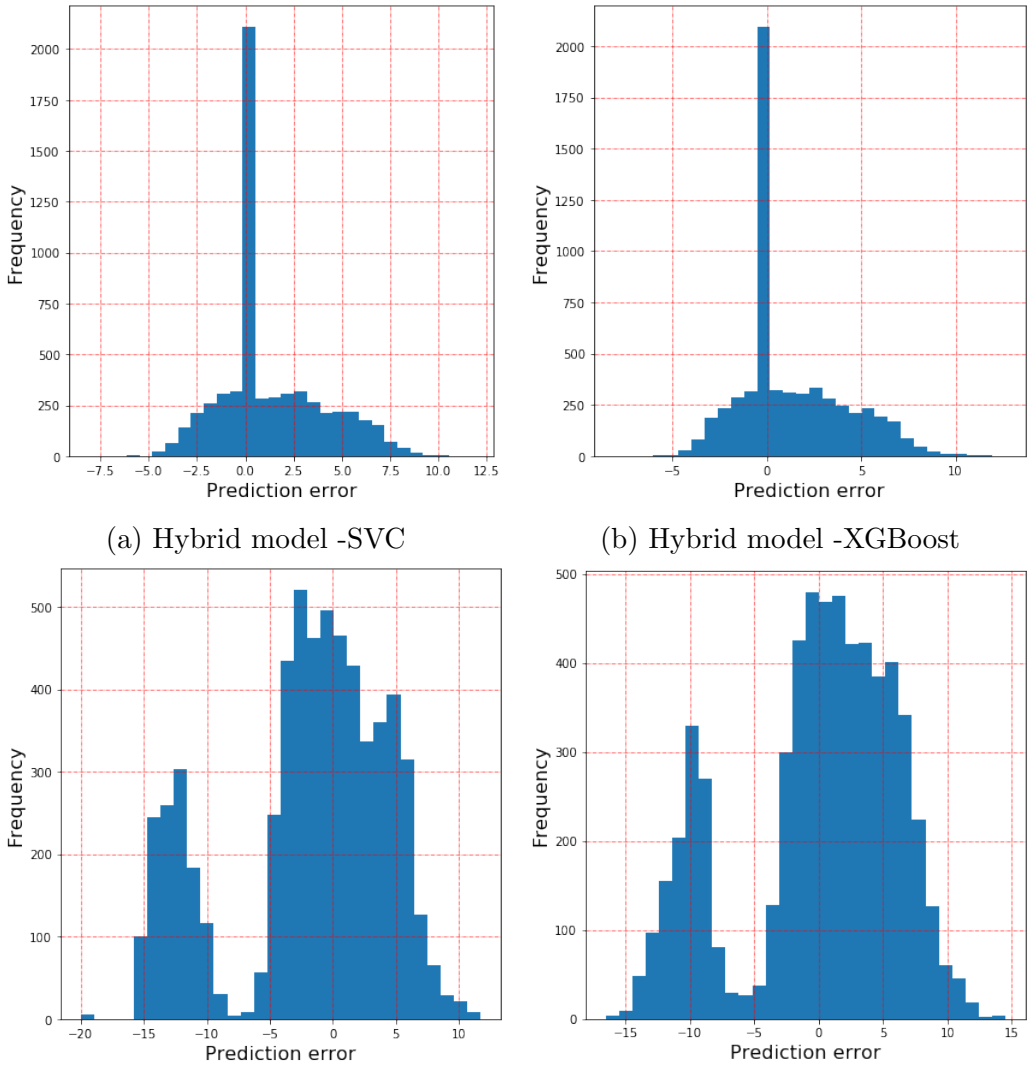
model based on the machine learning algorithm (XGBoost), the statistical model specifically for zero and N inflated model (mixture model) and a hybrid model with SVR and XGBoost classifier (Figure 7.1)

Table 7.1: Comparative results of all predictive models

	Pearson Cor. (Actual, Predicted)		**RMSE**	
	Train	Test	Train	Test
DT	0.259	0.264	6.133	6.205
RF	0.278	0.281	6.102	6.176
ANN	0.304	0.293	6.051	6.157
SVR	0.289	0.269	6.168	6.328
XGBoost	0.368	0.311	5.923	6.113
MLR	0.244	0.242	6.158	6.242
Binomial	0.244	0.242	6.158	6.242
ZNIB	0.091	0.057	6.613	6.754
Mixture model	0.225	0.208	6.456	6.607
Hybrid-RF	0.901	0.908	3.040	2.970
Hybrid-SVR	0.901	0.909	3.038	2.968
Hybrid-XGBoost	0.898	0.905	3.085	3.031

Comparative results of all predictive models (Table 7.1) and comparative prediction error plots for test data (Figure 7.1) clearly indicate that the developed hybrid model outperformed all other existing advanced predictive modeling techniques for inflated count data and can accurately predict the patent renewal life for Indian patent data.

The second objective of our research work was to investigate the important variables impacting the patent renewal life and thus the patent value, and patent quality.



(a) Hybrid model -SVC

(b) Hybrid model -XGBoost

(c) Mixture model

(d) XGBoost model

Figure 7.1: Comparative prediction error plots

We used permutation importance technique to extract the relative importance of variables from machine learning models. The top five important variable of few machine learning models are tabulated in Table 4.13. Comparative permutation importance

plots Figure 4.15 of all models and Table 4.13 clearly indicates that the grant lag and family size of the patent is most significant patent characteristics affecting patent renewal life. The few top important variables impacting patent renewal life are grant lag, family size, patentee type such as individual and institutional, ownership.

The modeling of zero and N inflated data poses a great challenge to any advanced techniques and it was clearly established during our research work. The advanced machine learning techniques as well as specifically statistical models for inflated count data could not perform efficiently for Indian patent data. Thus our research's third objective was to develop a more appropriate methodology to handle zero and N inflated binomial data distribution. We developed two methodologies — mixture model and the hybrid model. The developed mixture model was similar to the ZNIB model proposed by Sweeney, Haslett, and Parnell (2014) but it differs from their model as we assumed the source of excess data points are different for zero and any integer N. We further observed that the estimation of parameters in the statistical method was done through the EM algorithm due to computational challenges. We used the maximum likelihood method for the estimation of parameters using the python algopy package, which is basically algorithmic differentiation package. The estimated parameter values using standard package and our approach results in Table 5.1 for binomial regression are a close match and validate our pa-

rameter estimation approach. Although the mixture model performed at par with the existing ZNIB model still predictive model's efficiency scope still existed. We developed a hybrid model by harnessing the benefits of machine learning classification algorithms and binomial regression algorithms. The comparative results in Table 7.1 of predictive models indicate that the hybrid model outperformed all the existing models and predicts accurate patent renewal life. Our research also suggests that the mixture model as well as a hybrid model concept can be extended to other inflated common count data distribution for efficient predictive modeling.

7.2 Contributions

A noteworthy aspect of this study is the predictability of patent renewal lifespans, which ultimately provides insights into patent value, patent quality, and potential for future revenue generation. The developed patent renewal life predictive models can accurately predict the patent life span for Indian patents. Apart from the development of a predictive model, the study also contributed to the literature of patents through the identification of the important patent characteristics impacting Indian patent renewal life.

Our research work also contributed to the field of inflated count data modeling through the development of two new methodologies: the mixture model and the

hybrid model. A new approach for parameter estimation through the maximum likelihood method was developed during the development of the mixture model. The major contribution to inflated count data distribution predictive modeling was the development of hybrid models. The hybrid model not only crossed the barrier of all the advanced modeling techniques of handling zero and N binomial inflated data but also opened the path for the modeling of other inflated count data distributions such as inflated zero Poisson, inflated zero and N Poisson, zero and N inflated Binomial distribution. Hybrid model methodology can be extended to other inflated count data distributions based on predictive modeling..

7.3 Research implication

An accurate prediction of patent renewal life plays a significant role in the patent ecosystem as it serves as an indicator of a patent's value, quality, and potential for future revenue generation. Our research mainly delves into the accurate predictability of the patent renewal life of Indian patents. Developed predictive models can accurately predict the patent renewal life, which in turn can aid in assessing the patent value and quality. Additionally, our novel approaches for predictive models can handle inflated data sets. We can use the developed mixture and hybrid model to predict any zero or N-inflated count data distribution.

7.4 Practical implication

Patents are very critical for the innovation ecosystem and patent strategy plays a very crucial role not only for established companies but also for new startups. The predictive models developed in our research may accurately predict the renewal life of Indian patents. Our predictive model can provide the expected life of Indian patent life to the innovators and the companies and thus it can play a vital role in making informed decisions in research and development investment. An accurate prediction may be helpful in making a company strategy as a longer patent life may be a barrier for other competitors in the business while a shorter patent life prediction may be an indication of technological obsolescence and strong market competition. Based on the predictive life of patent companies even can make a decision on the investment in product development and product life cycle. Thus practically, the predictive model of Indian patents can be utilized in investment in research and development, strategy planning, and market dynamics.

On the one hand, patent transactions, which involve selling and licensing patents, help innovators monetize their research and bring the developed technology to benefit society; on the other hand, they help companies gain a competitive advantage and build their market position through patent acquisition. As patents are traded by companies in the market, indicators of patent values and patent qualities play an

essential role in patent acquisition decisions. Patent renewal life prediction can be used as an indicator of patent value and patent quality, thus helping firms make decisions on patent acquisition.

7.5 Limitations and Future Research

There are a few limitations in our study that may be considered in future work. One of the major limitations of the earlier study is that we considered only granted patents, which either expired or matured, and not the patents that are still alive. Further, our study was limited to Indian patent data. The proposed models can be utilized to model other countries' data sets, such as US patents and European patents.

Our research has not examined which patents contribute most to the inflated data points, specifically those with patent lives of zero and twenty years. It's important to conduct an investigation of the patents that significantly contribute to these inflated data points.

In our study, we split the original data into two sets, i.e., the train and test data sets. Further, we used RMSE and Pearson's correlation as criteria for comparison of model performance. Future work may consider splitting the original data into three sets, i.e., the train, test, and validation data sets, and other model performance

criteria such as hamming distance and gamma (a measure of rank correlation). Other advanced machine learning algorithms, such as deep neural networks, can also be considered in future work.

The proposed hybrid model was implemented for the inflated zero and N inflated binomial distributions; however, future studies can be done on other zero and N inflated count data distributions.

References

Allison, J. R., Lemley, M. A., Moore, K. A., & Trunkey, R. D. (2003). Valuable patents. *Geo. Lj*, *92*, 435.

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347.

Ambrammal, S. K., & Sharma. (2016). Impact of patenting on firms' performance: An empirical investigation based on manufacturing firms in india. *Economics of Innovation and New Technology*, *25*(1), 14–32.

Aristodemou, L., & Tietze, F. (2018). The state-of-the-art on intellectual property analytics (ipa): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (ip) data. *World Patent Information*, *55*, 37–51.

Arora, M., & Chaganty, N. R. (2021). Em estimation for zero-and k-inflated poisson regression model. *Computation*, *9*(9), 94.

Bass, S., & Kurgan, L. (2010). Discovery of factors influencing patent value based on machine learning in patents in the field of nanotechnology. *Scientometrics*, *82*(2), 217–241.

Baudry, M., & Dumont, B. (2006). Patent renewals as options: Improving the mechanism for weeding out lousy patents. *Review of Industrial Organization*, *28*, 41–62.

Bessen, J. (2008). The value of us patents by owner and patent characteristics. *Research Policy*, *37*(5), 932–945.

Bessen, J., & Meurer, M. J. (2008). Do patents perform like property? *Academy of Management Perspectives*, *22*(3), 8–20.

Bloom, N., & Reenen, J. (2002). Patents, real options and firm performance. *The Economic Journal*, *112*(478), C97–C116.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*, 144–152.

Breiman. (1984). Classification and regression trees. *The Wadsworth & Brooks/Cole*.

Breiman. (2001). Random forests. *Machine learning*, *45*, 5–32.

Cao, S., Zeng, Y., Yang, S., & Cao, S. (2021). Research on python data visualization technology. *Journal of Physics: Conference Series*, *1757*(1), 012122.

Chadha, A. (2009). Product cycles, innovation, and exports: A study of indian pharmaceuticals. *World Development*, *37*(9), 1478–1483.

Chen & Chang, R. (2021). Using machine learning to evaluate the influence of fintech patents: The case of taiwan's financial industry. *Journal of Computational and Applied Mathematics*, *390*, 113215.

Chen & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chi, Y.-C., & Wang, H.-C. (2022). Establish a patent risk prediction model for emerging technologies using deep learning and data augmentation. *Advanced Engineering Informatics*, *52*, 101509.

Choi, Jang, D., Jun, S., & Park, S. (2015). A predictive model of technology transfer using patent analysis. *Sustainability*, *7*(12), 16175–16195.

Choi, Jeong, B., Yoon, J., Coh, B.-Y., & Lee, J.-M. (2020). A novel approach to evaluating the business potential of intellectual properties: A machine learning-based predictive analysis of patent lifetime. *Computers & Industrial Engineering*, *145*, 106544.

Choi, Lee, H., Park, E. L., & Choi, S. (2019). Deep patent landscaping model using transformer and graph embedding. *arXiv preprint arXiv:1903.05823*.

Cohen, Nelson, R., & Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not).

Cohen Jr, A. C. (1960). Estimating the parameters of a modified poisson distribution. *Journal of the American Statistical Association*, *55*(289), 139–143.

Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27–36.

Danish, M. S., Ranjan, P., & Sharma, R. (2022). Assessing the impact of patent attributes on the value of discrete and complex innovations. *International Journal of Innovation Management*, *26*(02), 2250016.

Danish, M. S., Ranjan, P., & Sharma, R. (2021a). Determinants of patent survival in emerging economies: Evidence from residential patents in india. *Journal of Public Affairs*, *21*(2), e2211.

Danish, M. S., Ranjan, P., & Sharma, R. (2021b). *Identification of "valuable" technologies via patent statistics in india: An analysis based on renewal information* (tech. rep.). BASE University, Bengaluru, India.

Danish, M. S., Ranjan, P., & Sharma, R. (2020). Valuation of patents in emerging economies: A renewal model-based study of indian patents. *Technology analysis & strategic management*, *32*(4), 457–473.

Denton, F. R., & Heald, P. J. (2002). Random walks, non-cooperative games, and the complex mathematics of patent pricing. *Rutgers L. Rev.*, *55*, 1175.

Diop, A., Ba, D. B., & Lo, F. (2021). Asymptotic properties in the probit-zero-inflated binomial regression model. *arXiv preprint arXiv:2105.00483*.

Dunford, R., Su, Q., & Tamang, E. (2014). The pareto principle.

Dutt, R., Rathi, P., & Krishna, V. (2021). Novel mixed-encoding for forecasting patent grant duration. *World Patent Information*, *64*, 102007.

Ernst, H., Leptien, C., & Vitt, J. (2000). Inventors are not alike: The distribution of patenting output among industrial r&d personnel. *IEEE Transactions on engineering management*, *47*(2), 184–199.

Feller, W. (1943). On a general class of" contagious" distributions. *The Annals of mathematical statistics*, *14*(4), 389–400.

Fischer, T., & Leidinger, J. (2014). Testing patent value indicators on directly observed patent value—an empirical analysis of ocean tomo patent auctions. *Research policy*, *43*(3), 519–529.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*(417), 178–183.

Gulli, A., & Pal, S. (2017). *Deep learning with keras*. Packt Publishing Ltd.

Hall, D. B. (2000). Zero-inflated poisson and binomial regression with random effects: A case study. *Biometrics*, *56*(4), 1030–1039.

Han, E. J., & Sohn, S. Y. (2015). Patent valuation based on text mining and survival analysis. *The Journal of Technology Transfer*, *40*, 821–839.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research policy*, *32*(8), 1343–1363.

Harhoff, D., & Wagner, S. (2009). The duration of patent examination at the european patent office. *Management Science*, *55*(12), 1969–1984.

Hikkerova, Kammoun, N., & Lantz, J.-S. (2014). Patent life cycle: New evidence. *Technological Forecasting and Social Change*, *88*, 313–324.

Jaffe, A. (2015). Are patent fees effective at weeding out low-quality patents?

Kabore, F. P., & Park, W. G. (2019). Can patent family size and composition signal patent value? *Applied Economics*, *51*(60), 6476–6496.

Kim, M., & Geum, Y. (2020). Predicting patent transactions using patent-based machine learning techniques. *IEEE Access*, *8*, 188833–188843.

Kyebambe, M. N., Cheng, G., Huang, Y., He, C., & Zhang, Z. (2017). Forecasting emerging technologies: A supervised learning approach through patent analysis. *Technological Forecasting and Social Change*, *125*, 236–244.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14.

Lanjouw, J. O., Pakes, A., & Putnam, J. (1998). How to count patents and value intellectual property: The uses of patent renewal and application data. *The journal of industrial economics*, *46*(4), 405–432.

Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change*, *127*, 291–303.

Lemley, M. A., & Shapiro, C. (2005). Probabilistic patents. *Journal of Economic Perspectives*, *19*(2), 75–98.

Lerner, J. (1994). The importance of patent scope: An empirical analysis. *The RAND Journal of Economics*, 319–333.

Leung, T., & Sharma, R. (2021). Patenting in small and medium-sized enterprises: A systematic review and research agenda. *ournal of Business Research*, 202–216.

Marco, A. C., Sarnoff, J. D., & Charles, A. (2019). Patent claims and patent scope. *Research Policy*, *48*(9), 103790.

Maurseth, P. B. (2005). Lovely but dangerous: The impact of patent citations on patent renewal. *Economics of Innovation and New Technology*, *14*(5), 351–374.

Mirzamomen, Z., & Kangavari, M. R. (2017). A framework to induce more stable decision trees for pattern classification. *Pattern Analysis and Applications*, *20*, 991–1004.

Moore, K. A. (2005). Worthless patents. *Berkeley Technology Law Journal*, *20*(4), 1521–1552.

Neyman, J. (1939). On a new class of" contagious" distributions, applicable in entomology and bacteriology. *The Annals of Mathematical Statistics*, *10*(1), 35–57.

Nordhaus, W. D. (1969). Invention growth, and welfare: A theoretical treatment of technological change. *(No Title)*.

Pakes. (1984). *Patents as options: Some estimates of the value of holding european patent stocks* (tech. rep.). National Bureau of Economic Research.

Pakes & Schankerman, M. (1984). The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In *R&d, patents, and productivity* (pp. 73–88). University of Chicago Press.

Pakes, Simpson, M., Judd, K., & Mansfield, E. (1989). Patent renewal data. *Brookings papers on economic activity. Microeconomics*, *1989*, 331–410.

Poege, F., Harhoff, D., Gaessler, F., & Baruffaldi, S. (2019). Science quality and the value of inventions. *Science advances*, *5*(12), eaay7323.

Putnam. (1996). *The value of international patent rights*. Yale University.

Régibeau, P., & Rockett, K. (2010). Innovation cycles and learning at the patent office: Does the early patent get the delay? *The Journal of Industrial Economics*, *58*(2), 222–246.

Ridout, M., Demétrio, C. G., & Hinde, J. (1998). Models for count data with many zeros. *Proceedings of the XIXth international biometric conference*, *19*, 179–192.

Saint-Georges, M. d., & Potterie, B. v. P. d. l. (2013). A quality index for patent systems. *Research Policy*, *42*(3), 704–719.

Serrano, C. J. (2010). The dynamics of the transfer and renewal of patents. *The RAND Journal of Economics*, *41*(4), 686–708.

Sharma, R., Paswan, A. K., Ambrammal, S. K., & Dhanora, M. (2018). Impact of patent policy changes on r&d expenditure by industries in india. *The Journal of World Intellectual Property*, *21*(1-2), 52–69.

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*, 105524.

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on nuclear science*, *44*(3), 1464–1468.

Sullivan, R. J. (1994). Estimates of the value of patent rights in great britain and ireland, 1852-1876. *Economica*, 37–58.

Suominen, A., Toivanen, H., & Seppänen, M. (2017). Firms' knowledge profiles: Mapping patent data with unsupervised learning. *Technological Forecasting and Social Change*, *115*, 131–142.

Suzuki, J. (2011). Structural modeling of the value of patent. *Research Policy*, *40*(7), 986–1000.

Svensson, R. (2012). Commercialization, renewal, and quality of patents. *Economics of Innovation and New Technology*, *21*(2), 175–201.

Sweeney, J., Haslett, J., & Parnell, A. C. (2014). The zero & N-inflated binomial distribution with applications. *arXiv preprint arXiv:1407.0064*.

Tong, X., & Frame, J. D. (1994). Measuring national technological performance with patent claims data. *Research Policy*, *23*(2), 133–141.

Weber, B. (2014). Can safe ride programs reduce urban crime? *Regional Science and Urban Economics*, *48*, 1–11.

Yamada, H. (2022). Identification methods and indicators of important patents. *Library Hi Tech*, *40*(3), 750–785.

Zhu, F., Wang, X., Zhu, D., & Liu, Y. (2015). A supervised requirement-oriented patent classification scheme based on the combination of metadata and ci-

tation information. *International Journal of Computational Intelligence Systems*, *8*(3), 502–516.

Appendix A

Data collection process

The step-by-step procedure for data collection is discussed in this appendix. Recall that we had collected data on granted patents that were filed between 1995 and 2005. The following features were recorded for each patent: filing year, ownership, renewal year, number of claims, inventor size, family size, technology scope, grant lag, patentee type, and technology class.

1. Visit https://iprsearch.ipindia.gov.in/publicsearch

2. Enter the patent application number if you have a list of patents, or else mention the period for which you are extracting the data. Also tick the "Granted" box on top to fetch the list of granted patents.

3. Once you get the patent list, identify a patent (e.g., 1221/DEL/2000) on which you wish to collect the data.

   (a) Click on "E-Register" to easily obtain data on filing year, grant lag, renewal year, ownership, and patentee type.

   (b) Click on "Application Number" to get data on inventor size and classification (IPC). Schmoch **ipc** methodology was used here to identify the tech-

nology class as chemical, electrical, mechanical, instruments, and other fields.

(c) Click on "Application Status" and then "View Documents" to obtain data on number of claims.

4. Data on family size can be downloaded from PatSeer (https://patseer.com/), a private data source (license purchase is required).

Appendix B

Trends in Patent Applications

Year	2017-18	2018-19	2019-20	2020-21	2021-22
Filed	47854	50659	56267	58503	66440
Examined	60330	85426	80080	73165	66571
Granted	13045	15283	24936	28385	30073
Disposal	47695	50884	55945	52755	35990

Data source : Intellectual property India annual report 2021-22 (Page No.06) (https://ipindia.gov.in/writereaddata/Portal/Images/pdf/Final_Annual_Report_Eng_for_Net.pdf)

Notations

N	An integer > 0
k	An integer >0
p_i	Probability of success
m	Number of trials
L	Likelihood function
l	Log-likelihood function
$\sum$	Summation
$\prod$	Product
n_p	Number of predictors
X_i	Observations
g	Link function
η	linear predictor
α	Alpha parameter
β	Beta parameter
γ	Gamma parameter
ϕ	Probability
n	Total number of observations
n_1	Number of observations with renewal life value zero
n_2	Number of observations with renewal life value twenty
G	Gradient
H	Hessian
y_i	Observed values of patent renewal life
$\hat{y}_i$	Predicted values of patent renewal life
%	Percentage
ϵ	Error