

GerModel: Gestão de Modelos em Dados em Fluxo com Conceitos Recorrentes[‡]

Rodrigo Machado¹, Lucas Tavares¹, Laís Baroni¹,
Fabio Porto², Rafaelli Coutinho¹, Eduardo Ogasawara¹

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

²Laboratório Nacional de Computação Científica - LNCC

{rodrigo.machado, lucas.giusti, lais.baroni}@aluno.cefet-rj.br

fporto@lncc.br, rafaelli.coutinho@cefet-rj.br, eogasawara@ieee.org

Resumo. Fluxos meteorológicos de alta frequência apresentam não estacionariedade e concept drift recorrente, o que dificulta a manutenção de modelos preditivos. Este trabalho formula esse cenário como um problema de gestão de modelos em fluxo, no qual se decide entre retreinamento completo e reutilização de modelos históricos. Para isso, apresenta-se a GerModel, uma abordagem preliminar baseada no Population Stability Index (PSI) para selecionar e reutilizar modelos históricos em previsão horária de precipitação. A avaliação usa dados do aeroporto de Miami entre 2001 e 2025, com processamento mensal, e compara a GerModel com o retreinamento completo após mudança. No cenário avaliado, os resultados sustentam que a abordagem pode constituir alternativa viável, com F1 competitivo, alto reuso e menor custo computacional.

Abstract. High-frequency meteorological streams are non-stationary and subject to recurring concept drift, which makes predictive model maintenance difficult. This paper formulates this setting as a stream model management problem in which full retraining and historical model reuse must be balanced. To address this issue, we present GerModel, a preliminary PSI-based approach for selecting and reusing historical models in hourly precipitation forecasting. The evaluation uses Miami airport data from 2001 to 2025 with monthly processing and compares GerModel against full retraining after change detection. In the evaluated scenario, results indicate that the approach can be a viable alternative, with competitive F1, high reuse, and lower computational cost.

1. Introdução

Dados em fluxo são intrinsecamente não estacionários, com distribuições que evoluem ao longo do tempo [Read and Zliobaite, 2025]. Em diversos domínios, essa dinâmica inclui padrões sazonais e recorrentes, nos quais comportamentos previamente observados reaparecem, tornando o aprendizado contínuo um desafio [Anderson et al., 2019].

Nesse cenário, ocorre o *concept drift*, caracterizado por mudanças na distribuição dos dados ao longo do tempo, que podem afetar o desempenho de modelos preditivos

*Os autores agradecem à CAPES, FAPERJ e CNPq pelo apoio parcial à pesquisa.

[‡]Ferramentas de IA generativa são utilizadas apenas para revisão linguística. Os autores revisaram todas as sugestões e permanecem responsáveis pelo conteúdo do artigo.

[Gama et al., 2014]. Em contextos com recorrência, tais mudanças podem ser temporárias, permitindo reutilizar conceitos previamente aprendidos [Anderson et al., 2019].

Embora eficaz, o retreinamento completo após cada detecção de *concept drift* tende a ser custoso em cenários de alto volume e baixa latência [Gama et al., 2014]. Apesar dos avanços na detecção de *concept drift*, ainda há lacunas em mecanismos práticos de gestão de modelos em cenários com recorrência, especialmente no que diz respeito à decisão entre retreinamento e reutilização com base na similaridade entre distribuições.

Para enfrentar esse desafio, propõe-se a GerModel, uma abordagem que seleciona modelos armazenados por meio de similaridade distribucional, usando o *Population Stability Index* (PSI) como critério operacional para comparar o conceito corrente com conceitos históricos [Lu et al., 2025]. A métrica é adotada por sua eficácia em quantificar desvios na estabilidade dos atributos [Yurdakul and Naranjo, 2020], e sua adequação é avaliada por meio de métricas preditivas, eficiência computacional e taxa de reutilização.

A abordagem é avaliada no contexto de previsão horária de precipitação com dados do *Miami International Airport* (MIA), que apresentam contraste entre períodos chuvosos e secos, favorecendo a análise de recorrência em fluxo real [Misra and Jayasankar, 2024; Peeling et al., 2023].

A principal contribuição deste trabalho consiste em apresentar e avaliar a GerModel, comparando estratégias de reuso com o retreinamento completo sob diferentes mecanismos de acionamento, com dados reais de precipitação do aeroporto MIA no período de 2001 a 2025.

2. Fundamentação Teórica

Em aprendizado supervisionado, o objetivo é prever uma variável-alvo Y a partir de variáveis de entrada X . A distribuição conjunta $P(X, Y)$ reúne tanto a distribuição das entradas $P(X)$ quanto a distribuição da variável-alvo dado X , $P(Y|X)$ [Gama et al., 2014]. Quando essa distribuição muda ao longo do tempo, ocorre o *concept drift*, o que pode degradar o desempenho de modelos previamente ajustados [Gama et al., 2014].

Em séries meteorológicas, como as de precipitação horária, o *concept drift* pode ser abrupto, gradual, incremental ou recorrente [Gama et al., 2014; Webb et al., 2016]. A recorrência é particularmente relevante, pois padrões podem reaparecer e viabilizar o reuso de modelos associados a conceitos passados [Halstead et al., 2021].

O *concept drift* real, caracterizado por mudanças em $P(Y|X)$, distingue-se do virtual, associado a alterações em $P(X)$ sem impacto imediato na relação preditiva [Gama et al., 2014]. Essa distinção orienta estratégias de adaptação: passivas, nas quais a adaptação ocorre de forma sistemática, sem detecção explícita de mudanças; e ativas, nas quais a adaptação depende da detecção de mudanças estatísticas relevantes em $P(X)$ ou $P(Y|X)$ [Gama et al., 2014]. Em cenários com recorrência, contudo, o desafio vai além da detecção de mudanças e envolve a decisão de quando reutilizar ou atualizar modelos, considerando o compromisso entre custo computacional e desempenho preditivo.

Como critério distribucional, adota-se o PSI, definido por $PSI = \sum_{i=1}^n (P_i - Q_i) \ln \frac{P_i}{Q_i}$. Essa medida quantifica a divergência entre distribuições discretizadas, em que P_i e Q_i representam proporções nas distribuições de referência e observada, respectiva-

mente [Lu et al., 2025]. No contexto deste trabalho, o PSI é utilizado como sinal operacional de mudança no perfil de entrada e, portanto, como critério de similaridade para reuso de modelos. Embora eficiente, o PSI captura apenas mudanças em $P(X)$, desconsiderando $P(Y|X)$. Assim, seu uso na seleção de modelos deve ser complementado por métricas de desempenho preditivo [Lu et al., 2025; Yurdakul and Naranjo, 2020].

3. Trabalhos Relacionados

A literatura sobre aprendizado em fluxo com *concept drift* costuma distinguir abordagens voltadas à detecção de mudança daquelas voltadas à adaptação do modelo [Gama et al., 2014; Read and Zliobaite, 2025]. No primeiro grupo, destacam-se os detectores baseados em erro, como o *Drift Detection Method* (DDM), que monitoram a degradação do desempenho [Gama et al., 2014; Bayram et al., 2022]. Ainda no contexto de detecção, abordagens distribucionais acompanham alterações em $P(X)$ por meio de medidas de distância ou divergência e podem complementar detectores baseados em erro quando múltiplos indicadores são combinados [Read and Zliobaite, 2025; Tavares et al., 2025].

No segundo grupo, estratégias de adaptação atualizam o modelo após a detecção de mudança, geralmente por meio de retreinamento ou substituição do modelo corrente [Gama et al., 2014; Read and Zliobaite, 2025]. Apesar de consolidadas, essas abordagens nem sempre tratam a recorrência de conceitos como eixo central da gestão de modelos.

Quando a recorrência ocupa o centro da análise, o foco se desloca para abordagens voltadas à adaptação, com ênfase em mecanismos de memória, reativação e seleção de modelos. Métodos como o *Fast Adapting Ensemble* (FAE) [Ortíz Díaz et al., 2015], o *Recurring Adaptive Classifier Ensemble* (RACE) [Anderson et al., 2019], bem como outras abordagens de reuso e gerenciamento de memória, armazenam modelos e os reativam quando conceitos semelhantes reaparecem [Zhao et al., 2020; Halstead et al., 2021]. Outras propostas, como grafos de classificadores, modelam transições conceituais para orientar a escolha do modelo mais adequado [Sun et al., 2018]. Em comum, esses estudos tratam a recorrência como um problema de gestão de modelos, mas permanece em aberto a definição de critérios operacionais para selecionar modelos históricos. Nesse ponto, a GerModel se distingue por usar o PSI como regra operacional para decidir entre reuso e retreinamento completo.

4. Método

Esta seção descreve a GerModel, uma abordagem preliminar de gestão de modelos. Em termos operacionais, a GerModel decide, para cada conceito corrente, entre reutilizar um modelo previamente armazenado e executar um retreinamento completo com base em um limiar de similaridade e em restrições de memória.

A Figura 1 apresenta o fluxo completo do método. O diagrama destaca três etapas principais: (i) o bloco de acionamento, que verifica se há necessidade de adaptação; (ii) a lógica de seleção, que identifica o modelo histórico mais adequado ao reuso; e (iii) o bloco de decisão, que define entre reuso e retreinamento.

Quando o mecanismo de acionamento sinaliza adaptação, o sistema compara o conceito corrente C_t com um conjunto de modelos históricos $\mathcal{M} = \{M_1, \dots, M_k\}$, em que cada modelo M_k está associado ao *batch* de referência no qual foi treinado. Para cada

M_k , calcula-se o PSI médio $\psi(M_k, C_t) = \frac{1}{p} \sum_{j=1}^p \text{PSI}_j(M_k, C_t)$, em que p é o número de atributos considerados. Para a comparação distribucional via PSI, atributos numéricos são discretizados em faixas equiprováveis baseadas na distribuição de referência, enquanto atributos categóricos utilizam frequências relativas diretamente.

Define-se como candidato ao reuso o modelo histórico que minimiza a divergência PSI, $M^* = \arg \min_{M_k \in \mathcal{M}} \psi(M_k, C_t)$. Seja τ um limiar de similaridade definido pelo usuário (ou ajustado empiricamente). Se $\psi(M^*, C_t) \leq \tau$, reutiliza-se M^* ; caso contrário, executa-se retreinamento completo M_{novo} , incorporando o novo modelo a $\mathcal{M}$. Dessa forma, o PSI não substitui a avaliação preditiva, mas atua como critério operacional de triagem para a decisão de reuso.

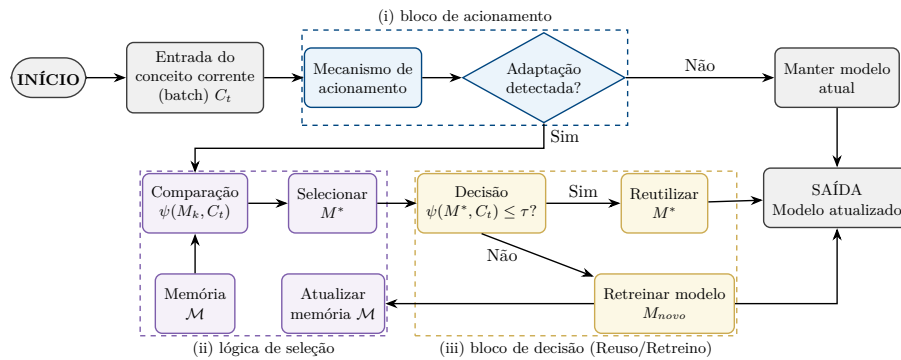


Figura 1. Fluxograma do método proposto: bloco de acionamento, lógica de seleção baseada em PSI e decisão entre reuso e retreinamento.

O método é implementado em R com `daltoolbox` e `heimdall` (CRAN-R) [Ogasawara et al., 2025a; Tavares et al., 2025]. O acionamento da adaptação é realizado com DDM ou com a estratégia PASSIVE. O DDM monitora a taxa de erro e seu desvio padrão ao longo de janelas, permitindo identificar mudanças abruptas ou graduais no desempenho [Gama et al., 2014; Bayram et al., 2022], enquanto a estratégia PASSIVE aciona a adaptação em todos os *batches*, representando uma condição de adaptação sistemática útil como contraste experimental [Ogasawara et al., 2025b; Tavares et al., 2025].

5. Avaliação Experimental

A avaliação experimental utiliza observações horárias do Aeroporto Internacional de Miami (MIA), obtidas via *Iowa Environmental Mesonet* [Iowa State University Mesonet, 2026]. O pré-processamento compreende a remoção de valores faltantes, a derivação de metadados temporais e a transformação da variável-alvo `p01i` em variável binária (precipitação $\geq 0,01$ polegadas). Para mitigar o desbalanceamento da classe positiva, aplica-se a técnica de *oversampling* (SMOTE) nos *batches* de treinamento, utilizando como preditores atributos climáticos que caracterizam o estado atmosférico: temperatura do ar (`tmpf`), ponto de orvalho (`dwpf`), umidade relativa (`relh`), direção do vento (`drct`) e velocidade do vento (`sknt`).

O fluxo de adaptação é executado mensalmente, tratando cada mês como um conceito corrente C_t em um fluxo contínuo [Read and Zliobaite, 2025]. Quatro classificadores, Naive Bayes (NB), Decision Tree (DT), Random Forest (RF) e Multi-Layer Perceptron (MLP), implementados via `daltoolbox`, são avaliados sob duas estratégias de acionamento: DDM, com parâmetros `min_instances = 2000`, `warning = 3`

e `out_control` = 4, e a abordagem PASSIVE [Gama et al., 2014; Bayram et al., 2022]. Como *baseline*, adota-se o retreinamento completo do modelo após cada acionamento de adaptação, sem reuso de modelos históricos. A decisão de reuso baseia-se no PSI, adotando-se $\tau = 0,25$ como limiar para mudanças distribucionais relevantes, valor usualmente associado a mudanças substanciais nas distribuições [Lu et al., 2025]. Para os atributos numéricos, utilizam-se 10 faixas equiprováveis definidas a partir da distribuição de referência, o que favorece a comparabilidade e a estabilidade das estimativas [Lu et al., 2025; Kurian and Allali, 2024].

O desempenho preditivo é avaliado por meio do F1-score médio e de seu desvio padrão, além de acurácia, precisão e *recall*. A eficiência computacional é medida pelo tempo gasto no processamento de cada *batch* mensal (detecção de *concept drift*, avaliação de reuso e treinamento). A taxa de reuso é definida por $\text{Reuso} = 1 - \frac{N_{\text{treinos}}}{N_{\text{drifts}}}$, em que N_{drifts} é o número de eventos de *concept drift* e N_{treinos} é o número de novos treinamentos realizados.

Para a análise estatística, os resultados mensais são agregados anualmente, utilizando o ano como unidade consolidada de análise. Sobre essas medidas, aplica-se o teste não paramétrico de Wilcoxon pareado, com ajuste de Bonferroni para controle de comparações múltiplas. Em cada comparação, reporta-se o *p_value*, adotando-se $\alpha = 0,05$ como critério de significância estatística. Quando $p_value \leq 0,05$, a diferença entre as estratégias é considerada significativa, e sua magnitude é complementada pelo *effect size* r , interpretado segundo a convenção de efeito pequeno, médio ou grande [Peres, 2026].

6. Resultados

A Tabela 1 apresenta os resultados de todas as configurações testadas, separadas pelo mecanismo de acionamento da adaptação. Considerando o F1 como métrica principal, o GerModel é competitivo em relação ao *baseline* e supera-o nas melhores configurações com MLP em ambos os mecanismos. Em DDM, o F1 aumenta de 39,3% no *baseline* para 42,4% com GerModel; em PASSIVE, sobe de 39,7% para 41,4%.

Além do ganho em F1, o GerModel reduz expressivamente o tempo de processamento. Em DDM, o tempo médio cai de 188 ms no *baseline* para 48 ms com GerModel; em PASSIVE, de 287 ms para 34 ms. Essas reduções estão associadas às taxas de reuso nas melhores configurações (98,0% em DDM e 98,7% em PASSIVE), indicando menor custo computacional sem perda no desempenho principal.

A Figura 2 resume o comportamento das demais configurações. Visualmente, MLP concentra os maiores valores centrais de F1 em ambos os mecanismos, enquanto DT, RF e NB permanecem em faixas inferiores e próximas entre si. Nas demais configurações, as diferenças entre GerModel e *baseline* são pequenas e, em geral, não significativas pelo teste de Wilcoxon, o que reforça a redução do tempo de processamento como principal benefício.

A superioridade das melhores configurações com MLP também é sustentada pelos testes estatísticos reportados na tabela. Em DDM, a melhor configuração com GerModel apresenta $p_W = 0,0013$ e magnitude grande ($r = 1,76$); em PASSIVE, os valores são $p_W = 0,0406$ e $r = 1,37$, também com magnitude grande. Assim, o ganho em F1 não é apenas numérico, mas também estatisticamente sustentado e relevante em termos

práticos.

Em termos comparativos, DDM favorece o melhor desempenho absoluto em F1, enquanto PASSIVE oferece o melhor compromisso entre eficiência e reuso, com o menor tempo médio de processamento. Em ambos os casos, a GerModel mantém ou melhora a capacidade preditiva em relação ao *baseline* e reduz substancialmente o custo computacional.

Tabela 1. Configurações ordenadas por F1 e separadas por mecanismo de acionamento da adaptação

Rank	Tipo	Class	F1(%)	Rec(%)	Prec(%)	Acc(%)	Reuso(%)	T _{proc}	p _W	W _{sig}	r	Mag _r
ACIONAMENTO: DDM												
1	GerModel	MLP	42,4	66,7	31,1	84,7	98,0	48	0,0013	SIG	1,76	G
4	baseline	MLP	39,3	74,5	26,7	80,7	0,0	188	–	–	–	–
5	baseline	DT	38,3	72,3	26,0	80,4	0,0	85	–	–	–	–
9	GerModel	DT	37,3	69,3	25,5	80,4	98,1	46	1,0000	NSIG	–	–
10	baseline	RF	37,2	38,5	35,9	89,1	0,0	117	–	–	–	–
11	GerModel	RF	36,7	35,2	38,3	89,9	98,1	50	1,0000	NSIG	–	–
13	GerModel	NB	36,6	77,7	23,9	77,4	98,0	324	0,0538	NSIG	–	–
16	baseline	NB	35,2	76,3	22,9	76,4	0,0	347	–	–	–	–
ACIONAMENTO: PASSIVE												
2	GerModel	MLP	41,4	67,9	29,7	83,8	98,7	34	0,0406	SIG	1,37	G
3	baseline	MLP	39,7	75,5	26,9	80,8	0,0	287	–	–	–	–
6	baseline	DT	37,6	72,7	25,3	79,8	0,0	83	–	–	–	–
7	baseline	RF	37,5	38,6	36,5	89,3	0,0	107	–	–	–	–
8	GerModel	DT	37,3	69,2	25,5	80,4	98,7	27	1,0000	NSIG	–	–
12	GerModel	RF	36,6	34,8	38,6	89,9	98,7	29	1,0000	NSIG	–	–
14	GerModel	NB	35,7	77,9	23,2	76,4	98,7	30	1,0000	NSIG	–	–
15	baseline	NB	35,3	78,1	22,8	76,0	0,0	74	–	–	–	–

F1: F1-score; Rec: recall; Prec: precisão; Acc: acurácia; Reuso: taxa de reuso (%); T_{proc}: tempo de processamento (ms)
p_W: p-value ajustado (Bonferroni), SIG: p_W < 0,05; r: effect size; Mag_r: magnitude do effect size r (P: pequeno, M: médio, G: grande).

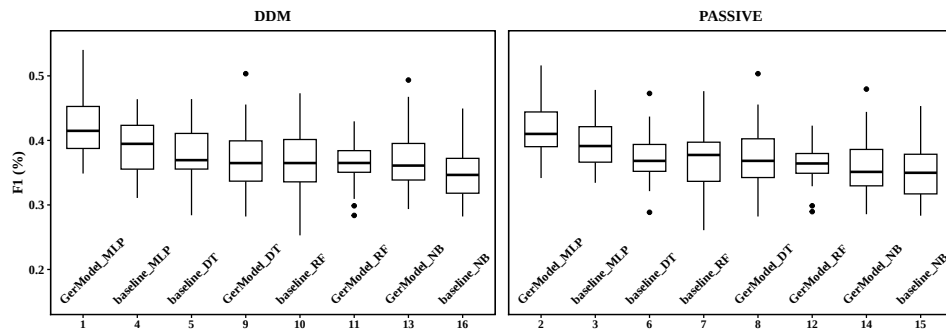


Figura 2. Distribuição anual do F1 por configuração, ordenada no eixo X pelo rank decrescente de F1 e separada por mecanismo de acionamento da adaptação.

7. Conclusão

Este trabalho investiga se a reutilização de modelos históricos pode constituir alternativa viável ao retreinamento completo em séries horárias de precipitação com recorrência de conceitos, usando dados reais do Aeroporto Internacional de Miami de 2001 a 2025. No cenário avaliado, os resultados indicam que a reutilização orientada por similaridade distribucional mantém desempenho competitivo e reduz o custo de adaptação. Como trabalho futuro, pretende-se ampliar a avaliação para outros domínios.

Referências

- Anderson, R., Koh, Y. S., Dobbie, G., and Bifet, A. (2019). Recurring concept meta-learning for evolving data streams. *Expert Systems with Applications*, 138:112832.
- Bayram, F., Ahmed, B. S., and Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245.
- Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46.
- Halstead, B., Koh, Y. S., Riddle, P., Pears, R., Pechenizkiy, M., and Bifet, A. (2021). Recurring concept memory management in data streams: exploiting data stream concept evolution to improve performance and transparency. *Data Mining and Knowledge Discovery*, 35:796 – 836.
- Iowa State University Mesonet (2026). FL ASOS Data Download. Technical report.
- Kurian, J. F. and Allali, M. (2024). Detecting drifts in data streams using Kullback-Leibler (KL) divergence measure for data engineering applications. *Journal of Data, Information and Management*, 6:207 – 216.
- Lu, S.-C., Song, W., Pfob, A., and Gibbons, C. (2025). Assessing the representativeness of large medical data using population stability index. *BMC Medical Research Methodology*, 25.
- Misra, V. and Jayasankar, C. (2024). Advancing the Seasonal Outlook of the Wet Seasons of Florida. *Weather and Forecasting*, 39:1751 – 1759.
- Ogasawara, E., Castro, A., Salles, D., Lima, J., Tavares, L., Carvalho, D., Bezerra, E., and Coutinho, R. (2025a). daltoolbox: Leveraging Experiment Lines to Data Analytics.
- Ogasawara, E., Salles, R., Porto, F., and Pacitti, E. (2025b). *Event Detection in Time Series*. Synthesis Lectures on Data Management. Springer Nature Switzerland, Cham, 1 edition.
- Ortíz Díaz, A., Del Campo-Ávila, J., Ramos-Jiménez, G., Frías Blanco, I., Caballero Mota, Y., Mustelíer Hechavarría, A., and Morales-Bueno, R. (2015). Fast adapting ensemble: A new algorithm for mining data streams with concept drift. *Scientific World Journal*, 2015:235810.
- Peeling, J. A., Judge, J., Misra, V., Jayasankar, C., and Lusher, W. R. (2023). Gap-free 16-year (2005–2020) sub-diurnal surface meteorological observations across Florida. *Scientific Data*, 10.
- Peres, F. F. (2026). Effect sizes for nonparametric tests. *Biochemia Medica*, 36.
- Read, J. and Zliobaite, I. (2025). Supervised Learning from Data Streams: An Overview and Update. *ACM Computing Surveys*, 57:1–37.
- Sun, Y., Wang, Z., Bai, Y., Dai, H., and Nahavandi, S. (2018). A Classifier Graph Based Recurring Concept Detection and Prediction Approach. *Computational Intelligence and Neuroscience*, 2018:4276291.
- Tavares, L., Carvalho, L., Machado, R., Carvalho, D., Pacitti, E., Porto, F., and Ogasawara, E. (2025). heimdall: Drift Adaptable Models.
- Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., and Petitjean, F. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery*, 30:964 – 994.
- Yurdakul, B. and Naranjo, J. (2020). Statistical properties of the population stability index. *Journal of Risk Model Validation*, 14:89 – 100.
- Zhao, P., Cai, L.-W., and Zhou, Z.-H. (2020). Handling concept drift via model reuse. *Machine Learning*, 109:533 – 568.